(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 February 2003 (20.02.2003)

PCT

(10) International Publication Number
WO 03/014322 A2

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): AZIMZAI, Yalda [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). BARROSO, Ines [PT/GB]; 38 Eden Street, Cambridge, Kent CB1 1EL (GB). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). BECHA, Shanya, D. [US/US]; 21062 Gary Drive # 117, Castro Valley, CA 94546 (US). BOROWSKY, Mark, L. [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). DUGGAN, Brendan, M. [AU/US]; 243 Buena Vista Avenue # 306, Sunnyvale, CA 94086 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). EMERLING, Brooke, M. [US/US]; 1735 Woodland Avenue # 71, Palo Alto, CA 94303 (US). FORSYTHE, Ian, J. [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). GIETZEN, Kimberly, J. [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). GORVAD, Ann, E. [US/US]; 369 Marie Common, Livermore, CA 94550 (US). GRAUL, Richard, C. [US/US]; 682-29th Avenue, San Francisco, CA 94121 (US). GRIFFIN, Jennifer, A. [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). GURURAJAN, Rajagopal [IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US). HAFALIA, April, J.A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). ISON, Craig, H. [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). KABLE, Amy, E. [US/US]; 2345 Polk Street #4, San Francisco, CA 94109 (US). KHAN, Farrah, A. [IN/US]; 3617 Central Road #102, Glenview, IL 60025 (US). LEE, Sally [US/US]; 3643 26th Street, San Francisco, CA 94110 (US). LEE, Soo Yeun [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). LI, Joana, X. [US/US]; 1264 Geneva Avenue, San Francisco, CA 94112 (US). REDDY, Roopa [IN/US]; 1233 West McKinley Drive # 3, Sunnyvale, CA 94086 (US). RICHARDSON, Thomas, W. [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). SPRAGUE, William, W. [US/US]; 611 13th Street # C, Sacramento, CA 95814 (US). SWARNAKAR, Anita [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). TANG, Y. Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). WARREN, Bridget, A. [US/US]; 1810 S. El Camino Real #B103, Encinitas, CA 94024 (US). XU, Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). YAO, Monique, G. [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). YUE, Huibin [US/US]; 1170 South Stelling Road, Cupertino, CA 95014 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(54) Title: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH

(57) Abstract: Various embodiments of the invention provide human proteins associated with cell growth, differentiation, and death (CGDD) and polynucleotides which identify and encode CGDD. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of CGDD.

GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH

TECHNICAL FIELD

The invention relates to novel nucleic acids, proteins associated with cell growth,

5   differentiation, and death encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of cell proliferative disorders including cancer, developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and proteins associated with cell growth,

10   differentiation, and death.

BACKGROUND OF THE INVENTION

Human growth and development requires the spatial and temporal regulation of cell differentiation, cell proliferation, and apoptosis. These processes coordinately control reproduction,

15   aging, embryogenesis, morphogenesis, organogenesis, and tissue repair and maintenance. At the cellular level, growth and development is governed by the cell's decision to enter into or exit from the cell division cycle and by the cell's commitment to a terminally differentiated state. These decisions are made by the cell in response to extracellular signals and other environmental cues it receives. The following discussion focuses on the molecular mechanisms of cell division,

20   embryogenesis, cell differentiation and proliferation, and apoptosis, as well as disease states such as cancer which can result from disruption of these mechanisms.

Cell Cycle

Cell division is the fundamental process by which all living things grow and reproduce. In unicellular organisms such as yeast and bacteria, each cell division doubles the number of organisms.

25   In multicellular species many rounds of cell division are required to replace cells lost by wear or by programmed cell death, and for cell differentiation to produce a new tissue or organ. Progression through the cell cycle is governed by the intricate interactions of protein complexes. This regulation depends upon the appropriate expression of proteins which control cell cycle progression in response to extracellular signals, such as growth factors and other mitogens, and intracellular cues, such as

30   DNA damage or nutrient starvation. Molecules which directly or indirectly modulate cell cycle progression fall into several categories, including cyclins, cyclin-dependent protein kinases, growth factors and their receptors, second messenger and signal transduction proteins, oncogene products, and tumor-suppressor proteins.

Details of the cell division cycle may vary, but the basic process consists of three principle

35   events. The first event, interphase, involves preparations for cell division, replication of the DNA,

and production of essential proteins. In the second event, mitosis, the nuclear material is divided and separates to opposite sides of the cell. The final event, cytokinesis, is division and fission of the cell cytoplasm. The sequence and timing of cell cycle transitions is under the control of the cell cycle regulation system which controls the process by positive or negative regulatory circuits at various

5   check points.

Mitosis marks the end of interphase and concludes with the onset of cytokinesis. There are four stages in mitosis, occurring in the following order: prophase, metaphase, anaphase and telophase. Prophase includes the formation of bi-polar mitotic spindles, composed of microtubules and associated proteins such as dynein, which originate from polar mitotic centers. During

10   metaphase, the nuclear material condenses and develops kinetochore fibers which aid in its physical attachment to the mitotic spindles. The ensuing movement of the nuclear material to opposite poles along the mitotic spindles occurs during anaphase. Telophase includes the disappearance of the mitotic spindles and kinetochore fibers from the nuclear material. Mitosis depends on the interaction of numerous proteins. For example, centromere-associated proteins such as CENP-A, -B, and -C,

15   play structural roles in kinetochore formation and assembly (Saffery, R. et al. (2000) Human Mol. Gen. 9:175-185).

During the M phase of eukaryotic cell cycling, structural rearrangements occur ensuring appropriate distribution of cellular components between daughter cells. Breakdown of interphase structures into smaller subunits is common. The nuclear envelope breaks into vesicles, and nuclear

20   lamins are disassembled. Subsequent phosphorylation of these lamins occurs and is maintained until telophase, at which time the nuclear lamina structure is reformed. cDNAs responsible for encoding M phase phosphorylation (MPPs) are components of U3 small nucleolar ribonucleoprotein (snoRNP), and relocalize to the nucleolus once mitosis is complete (Westendorf, J.M. et al. (1998) J. Biol. Chem. 9:437-449). U3 snoRNPs are essential mediators of RNA processing events.

25   Proteins involved in the regulation of cellular processes such as mitosis include the Ser/Thr-protein phosphatases type 1 (PP-1). PP-1s act by dephosphorylation of key proteins involved in the metaphase-anaphase transition. The gene PP1R7 encodes the regulatory polypeptide sds22, having at least six splice variants (Ceulemans, H. et al. (1999) Eur. J. Biochem. 262:36-42). Sds22 modulates the activity of the catalytic subunit of PP-1s, and enhances the PP-1-dependent dephosphorylation of

30   mitotic substrates.

Cell cycle regulatory proteins play an important role in cell proliferation and cancer. For example, failures in the proper execution and timing of cell cycle events can lead to chromosome segregation defects resulting in aneuploidy or polyploidy. This genomic instability is characteristic of transformed cells (Luca, F.C. and M. Winey (1998) Mol. Biol. Cell. 9:29-46). A recently

35   identified protein, mMOB1, is the mammalian homolog of yeast MOB1, an essential yeast gene

required for completion of mitosis and maintenance of ploidy. The mammalian mMOB1 is a member of protein complexes including protein phosphatase 2A (PP2A), and its phosphorylation appears to be regulated by PP2A (Moreno, C.S. et al. (2001) J. Biol. Chem. 276:24253-24260). PP2A has been implicated in the development of human cancers, including lung and colon cancers and leukemias.

5   ERM proteins are responsible for the cross-linking of actin filaments to the plasma membrane. FERM domains, located at the N-terminal regions of ERM proteins, regulate interactions between the cytoplasmic domains of the integrated membrane proteins with the membrane itself. The Protein 4.1 family of molecules are responsible for linking the actin cytoskeleton to cell surface glycoproteins. For example, the neurofibromatosis 2 (NF2) tumor suppressor is a member of the

10 Protein 4.1 family. NF2 proteins participate in suppression of cell growth, and retard other cytoskeletal-dependent functions including cell spreading, attachment and motility (Gutmann, D.H. et al. (2001) Neurobiol. Dis. 8:266-278). Recently, a novel putative tumor suppressor gene and member of the NF2/ERM/4.1 superfamily has been observed to retard the growth of non-small cell lung carcinoma cells (Tran, Y.K. et al. (1999) Cancer Res. 59:35-43).

15   Cell cycle regulation involves numerous proteins interacting in a sequential manner. The eukaryotic cell cycle consists of several highly controlled events whose precise order ensures successful DNA replication and cell division. Cells maintain the order of these events by making later events dependent on the successful completion of earlier events. This dependency is enforced by cellular mechanisms called checkpoints. Examples of additional cell cycle regulatory proteins

20 include the histone deacetylases (HDACs). HDACs are involved in cell cycle regulation, and modulate chromatin structure. Human HDAC1 has been found to interact *in vitro* with the human Hus1 gene product, whose *Schizosaccharomyces pombe* homolog has been implicated in G_2/M checkpoint control (Cai, R.L. et al. (2000) J. Biol. Chem. 275:27909-27916).

        DNA damage (G_2) and DNA replication (S-phase) checkpoints arrest eukaryotic cells at the

25 G_2/M transition. This arrest provides time for DNA repair or DNA replication to occur before entry into mitosis. Thus, the G_2/M checkpoint ensures that mitosis only occurs upon completion of DNA replication and in the absence of chromosomal damage. The Hus1 gene of *Schizosaccharomyces pombe* is a cell cycle checkpoint gene, as are the rad family of genes (e.g., rad1 and rad9) (Volkmer, E. and L.M. Karnitz (1999) J. Biol. Chem. 274:567-570; Kostrub C.F. et al. (1998) EMBO J.

30 17:2055-2066). These genes are involved in the mitotic checkpoint, and are induced by either DNA damage or blockage of replication. Induction of DNA damage or replication block leads to loss of function of the Hus1 gene and subsequent cell death. Human homologs have been identified for most of the rad genes, including ATM and ATR, the human homologs of rad3p. Mutations in the ATM gene are correlated with the severe congenital disease ataxia-telagiectasia (Savitsky, K. et al. (1995)

35 Science 268:1749-1753). The human Hus1 protein has been shown to act in a complex with rad1

protein which interacts with rad9, making them central components of a DNA damage-responsive protein complex of human cells (Volkmer and Karnitz, *supra*).

The entry and exit of a cell from mitosis is regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins act by binding to and activating a group of

5 cyclin-dependent protein kinases (Cdks) which then phosphorylate and activate selected proteins involved in the mitotic process. Cyclins are characterized by a large region of shared homology that is approximately 180 amino acids in length and referred to as the "cyclin box" (Chapman, D.L. and D.J. Wolgemuth (1993) Development 118:229-240). In addition, cyclins contain a conserved 9 amino acid sequence in the N-terminal region of the molecule called the "destruction box." This sequence is

10 believed to be a recognition code that triggers ubiquitin-mediated degradation of cyclin B (Hunt, T. (1991) Nature 349:100-101). Several types of cyclins exist (Ciechanover, A. (1994) Cell 79:13-21). Progression through G1 and S phase is driven by the G1 cyclins and their catalytic subunits, including Cdk2–cyclin A, Cdk2–cyclin E, Cdk4–cyclin D and Cdk6–cyclin D. Progression through the G2–M transition is driven by the activation of mitotic CDK–cyclin complexes such as Cdc2–cyclin A,

15 Cdc2–cyclin B1 and Cdc2–cyclin B2 complexes (reviewed in Yang, J. and S. Kornbluth (1999) Trends Cell Biol. 9:207-210).

Cyclins are degraded through the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins in eukaroytic cells and in some bacteria. The UCS mediates the elimination of abnormal proteins and regulates the half-lives of important regulatory proteins that

20 control cellular processes such as gene transcription and cell cycle progression. The UCS is implicated in the degradation of mitotic cyclin kinases, oncoproteins, tumor suppressor genes such as p53, viral proteins, cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, *supra*).

The process of ubiquitin conjugation and protein degradation occurs in five principle steps

25 (Jentsch, S. (1992) Annu. Rev. Genet. 26:179-207). First ubiquitin (Ub), a small, heat stable protein is activated by a ubiquitin-activating enzyme (E1) in an ATP dependent reaction which binds the C-terminus of Ub to the thiol group of an internal cysteine residue in E1. Second, activated Ub is transferred to one of several Ub-conjugating enzymes (E2). Different ubiquitin-dependent proteolytic pathways employ structurally similar, but distinct ubiquitin-conjugating enzymes that are

30 associated with recognition subunits which direct them to proteins carrying a particular degradation signal. Third, E2 transfers the Ub molecule through its C-terminal glycine to a member of the ubiquitin-protein ligase family, E3. Fourth, E3 transfers the Ub molecule to the target protein. Additional Ub molecules may be added to the target protein forming a multi-Ub chain structure. Fifth, the ubiquinated protein is then recognized and degraded by the proteasome, a large,

35 multisubunit proteolytic enzyme complex, and Ub is released for re-utilization.

Prior to activation, Ub is usually expressed as a fusion protein composed of an N-terminal ubiquitin and a C-terminal extension protein (CEP) or as a polyubiquitin protein with Ub monomers. attached head to tail. CEPs have characteristics of a variety of regulatory proteins; most are highly basic, contain up to 30% lysine and arginine residues, and have nucleic acid-binding domains

5   (Monia, B.P. et al. (1989) J. Biol. Chem. 264:4093-4103). The fusion protein is an important intermediate which appears to mediate co-regulation of the cell's translational and protein degradation activities, as well as localization of the inactive enzyme to specific cellular sites. Once delivered, C-terminal hydrolases cleave the fusion protein to release a functional Ub (Monia et al., *supra*).

10   Ub-conjugating enzymes (E2s) are important for substrate specificity in different UCS pathways. All E2s have a conserved domain of approximately 16 kDa called the UBC domain that is at least 35% identical in all E2s and contains a centrally located cysteine residue required for ubiquitin-enzyme thiolester formation (Jentsch, *supra*). A well conserved proline-rich element is located N-terminal to the active cysteine residue. Structural variations beyond this conserved domain

15   are used to classify the E2 enzymes. Class I E2s consist almost exclusively of the conserved UBC domain. Class II E2s have various unrelated C-terminal extensions that contribute to substrate specificity and cellular localization. Class III E2s have unique N-terminal extensions which are believed to be involved in enzyme regulation or substrate specificity.

The E2s are important for substrate specificity in several UCS pathways. All E2s have a

20   conserved ubiquitin conjugation (UBC) domain of approximately 16 kD, at least 35% identity with each other, and contain a centrally located cysteine residue which is necessary for ubiquitin-enzyme thiolester formation (Jentsch, supra). A highly conserved proline-rich element is located N-terminal to the active cysteine residue. Structural variations outside of this conserved domain are used to separate the E2 enzymes into classes. The E2s of class 1 (E2-1) consist of the conserved UBC

25   domain and include yeast E2-1 and UBCs 4, 5, and 7. These E2s are thought to require E3 to carry out their activities (Jentsch, supra). UBC7 has been shown to recognize ubiquitin as a substrate and to form polyubiquitin chains in vitro (van Nocker, S. et al. (1996) J. Biol. Chem. 271:12150-58). E2s of class 2 (E2-2) have various unrelated C-terminal extensions that contribute to substrate specificity and cellular localization. The yeast E2-2 enzymes, UBC2 and UBC3, have highly acidic C-terminal

30   extensions that promote interactions with basic substrates such as histones. Yeast UBC6 has a hydrophobic signal-anchor sequence that localizes the protein to the endoplasmic reticulum.

A mitotic cyclin-specific E2 (E2-C) is characterized by the conserved UBC domain, an N-terminal extension of 30 amino acids not found in other E2s, and a 7 amino acid unique sequence adjacent to this extension. These characteristics together with the high affinity of E2-C for cyclin

35   identify it as a new class of E2 (Aristarkhov, A. et al. (1996) Proc. Natl. Acad. Sci. 93:4294-99).

Ubiquitin-protein ligases (E3s) catalyze the last step in the ubiquitin conjugation process, covalent attachment of ubiquitin to the substrate. E3 plays a key role in determining the specificity of the process. Only a few E3s have been identified so far. One type of E3 ligases is the HECT (homologous to E6-AP C-terminus) domain protein family. One member of the family, E6-AP

5   (E6-associated protein) is required, along with the human papillomavirus (HPV) E6 oncoprotein, for the ubiquitination and degradation of p53 (Scheffner, M. et al. (1993) Cell 75:495-505). The C-terminal domain of HECT proteins contains the highly conserved ubiquitin-binding cysteine residue. The N-terminal region of the various HECT proteins is variable and is believed to be involved in specific substrate recognition (Huibregtse, J.M. et al. (1997) Proc. Natl Acad. Sci. USA

10   94:3656-3661). The SCF (Skp1-Cdc53/Cullin-F box receptor) family of proteins comprise another group of ubiquitin ligases (Deshaies, R. (1999) Annu. Rev. Dev. Biol. 15:435-467). Multiple proteins are recruited into the SCF complex, including Skp1, cullin, and an F box domain containing protein. The F box protein binds the substrate for the ubiquitination reaction and may play roles in determining substrate specificity and orienting the substrate for reaction. Skp1 interacts with both the

15   F box protein and cullin and may be involved in positioning the F box protein and cullin in the complex for transfer of ubiquitin from the E2 enzyme to the protein substrate. Substrates of SCF ligases include proteins involved in regulation of CDK activity, activation of transcription, signal transduction, assembly of kinetochores, and DNA replication.

Sgt1 was identified in a screen for genes in yeast that suppress defects in kinetochore

20   function caused by mutations in Skp1 (Kitagawa, K. et al. (1999) Mol. Cell 4:21-33). Sgt1 interacts with Skp1 and associates with SCF ubiquitin ligase. Defects in Sgt1 cause arrest of cells at either G1 or G2 stages of the cell cycle. A yeast Sgt1 null mutant can be rescued by human Sgt1, an indication of the conservation of Sgt1 function across species. Sgt1 is required for assembly of kinetochore complexes in yeast.

25   Abnormal activities of the UCS are implicated in a number of diseases and disorders. These include, e.g., cachexia (Llovera, M. et al. (1995) Int. J. Cancer 61:138-141), degradation of the tumor-suppressor protein, p53 (Ciechanover, *supra*), and neurodegeneration such as observed in Alzheimer's disease (Gregori, L. et al. (1994) Biochem. Biophys. Res. Commun. 203:1731-1738). Since ubiquitin conjugation is a rate-limiting step in antigen presentation, the ubiquitin degradation

30   pathway may also have a critical role in the immune response (Grant, E.P. et al. (1995) J. Immunol. 155:3750-3758).

Certain cell proliferation disorders can be identified by changes in the protein complexes that normally control progression through the cell cycle. A primary treatment strategy involves reestablishing control over cell cycle progression by manipulation of the proteins involved in cell

35   cycle regulation (Nigg, E.A. (1995) BioEssays 17:471-480).

6

Regulated progression of the cell cycle depends on the integration of growth control pathways with the basic cell cycle machinery. Cell cycle regulators have been identified by selecting for human and yeast cDNAs that block or activate cell cycle arrest signals in the yeast mating pheromone pathway when they are overexpressed. Known regulators include human CPR (cell cycle

5   progression restoration) genes, such as CPR8 and CPR2, and yeast CDC (cell division control) genes, including CDC91, that block the arrest signals. The CPR genes express a variety of proteins including cyclins, tumor suppressor binding proteins, chaperones, transcription factors, translation factors, and RNA-binding proteins (Edwards, M.C. et al.(1997) Genetics 147:1063-1076).

Several cell cycle transitions, including the entry and exit of a cell from mitosis, are

10   dependent upon the activation and inhibition of cyclin-dependent kinases (Cdks). The Cdks are composed of a kinase subunit, Cdk, and an activating subunit, cyclin, in a complex that is subject to many levels of regulation. There appears to be a single Cdk in Saccharomyces cerevisiae and Saccharomyces pombe whereas mammals have a variety of specialized Cdks. Cyclins act by binding to and activating cyclin-dependent protein kinases which then phosphorylate and activate selected

15   proteins involved in the mitotic process. The Cdk-cyclin complex is both positively and negatively regulated by phosphorylation, and by targeted degradation involving molecules such as CDC4 and CDC53. In addition, Cdks are further regulated by binding to inhibitors and other proteins such as Suc1 that modify their specificity or accessibility to regulators (Patra, D. and W.G. Dunphy (1996) Genes Dev. 10:1503-1515; and Mathias, N. et al. (1996) Mol. Cell Biol. 16:6634-6643).

20   **Reproduction**

The male and female reproductive systems are complex and involve many aspects of growth and development. The anatomy and physiology of the male and female reproductive systems are reviewed in (Guyton, A.C. (1991) Textbook of Medical Physiology, W.B. Saunders Co., Philadelphia PA, pp. 899-928).

25   The male reproductive system includes the process of spermatogenesis, in which the sperm are formed, and male reproductive functions are regulated by various hormones and their effects on accessory sexual organs, cellular metabolism, growth, and other bodily functions.

Spermatogenesis begins at puberty as a result of stimulation by gonadotropic hormones released from the anterior pituitary. Immature sperm (spermatogonia) undergo several mitotic cell

30   divisions before undergoing meiosis and full maturation. The testes secrete several male sex hormones, the most abundant being testosterone, that is essential for growth and division of the immature sperm, and for the masculine characteristics of the male body. Three other male sex hormones, gonadotropin-releasing hormone (GnRH), luteinizing hormone (LH), and follicle-stimulating hormone (FSH) control sexual function.

35   The uterus, ovaries, fallopian tubes, vagina, and breasts comprise the female reproductive

system. The ovaries and uterus are the source of ova and the location of fetal development, respectively. The fallopian tubes and vagina are accessory organs attached to the top and bottom of the uterus, respectively. Both the uterus and ovaries have additional roles in the development and loss of reproductive capability during a female's lifetime. The primary role of the breasts is lactation.

5   Multiple endocrine signals from the ovaries, uterus, pituitary, hypothalamus, adrenal glands, and other tissues coordinate reproduction and lactation. These signals vary during the monthly menstruation cycle and during the female's lifetime. Similarly, the sensitivity of reproductive organs to these endocrine signals varies during the female's lifetime.

        A combination of positive and negative feedback to the ovaries, pituitary and hypothalamus

10   glands controls physiologic changes during the monthly ovulation and endometrial cycles. The anterior pituitary secretes two major gonadotropin hormones, follicle-stimulating hormone (FSH) and luteinizing hormone (LH), regulated by negative feedback of steroids, most notably by ovarian estradiol. If fertilization does not occur, estrogen and progesterone levels decrease. This sudden reduction of the ovarian hormones leads to menstruation, the desquamation of the endometrium.

15   Hormones further govern all the steps of pregnancy, parturition, lactation, and menopause. During pregnancy large quantities of human chorionic gonadotropin (hCG), estrogens, progesterone, and human chorionic somatomammotropin (hCS) are formed by the placenta. hCG, a glycoprotein similar to luteinizing hormone, stimulates the corpus luteum to continue producing more progesterone and estrogens, rather than to involute as occurs if the ovum is not fertilized. hCS is similar to growth

20   hormone and is crucial for fetal nutrition.

        The female breast also matures during pregnancy. Large amounts of estrogen secreted by the placenta trigger growth and branching of the breast milk ductal system while lactation is initiated by the secretion of prolactin by the pituitary gland.

        Parturition involves several hormonal changes that increase uterine contractility toward the

25   end of pregnancy, as follows. The levels of estrogens increase more than those of progesterone. Oxytocin is secreted by the neurohypophysis. Concomitantly, uterine sensitivity to oxytocin increases. The fetus itself secretes oxytocin, cortisol (from adrenal glands), and prostaglandins.

        Menopause occurs when most of the ovarian follicles have degenerated. The ovary then produces less estradiol, reducing the negative feedback on the pituitary and hypothalamus glands.

30   Mean levels of circulating FSH and LH increase, even as ovulatory cycles continue. Therefore, the ovary is less responsive to gonadotropins, and there is an increase in the time between menstrual cycles. Consequently, menstrual bleeding ceases and reproductive capability ends.

Embryogenesis

        Mammalian embryogenesis is a process which encompasses the first few weeks of

35   development following conception. During this period, embryogenesis proceeds from a single

8

fertilized egg to the formation of the three embryonic tissues, then to an embryo which has most of its internal organs and all of its external features.

The normal course of mammalian embryogenesis depends on the correct temporal and spatial regulation of a large number of genes and tissues. These regulation processes have been intensely studied in mouse. An essential process that is still poorly understood is the activation of the embryonic genome after fertilization. As mouse oocytes grow, they accumulate transcripts that are either translated directly into proteins or stored for later activation by regulated polyadenylation. During subsequent meiotic maturation and ovulation, the maternal genome is transcriptionally inert, and most maternal transcripts are deadenylated and/or degraded prior to, or together with, the activation of the zygotic genes at the two-cell stage (Stutz, A. et al. (1998) Genes Dev. 12:2535-2548). The maternal to embryonic transition involves the degradation of oocyte, but not zygotic transcripts, the activation of the embryonic genome, and the induction of cell cycle progression to accommodate early development.

MATER (Maternal Antigen That Embryos Require) was initially identified as a target of antibodies from mice with ovarian immunity (Tong, Z-B. and L.M. Nelson (1999) Endocrinology 140:3720-3726). Expression of the gene encoding MATER is restricted to the oocyte, making it one of a limited number of known maternal-effect genes in mammals (Tong, Z-B. et al. (2000) Mamm. Genome 11:281-287). The MATER protein is required for embryonic development beyond two cells, based upon preliminary results from mice in which this gene has been inactivated. The 1111-amino acid MATER protein contains a hydrophilic repeat region in the amino terminus, and a region containing 14 leucine-rich repeats in the carboxyl terminus. These repeats resemble the sequence found in porcine ribonuclease inhibitor that is critical for protein-protein interactions.

The degradation of maternal transcripts during meiotic maturation and ovulation may involve the activation of a ribonuclease just prior to ovulation. Thus the function of MATER may be to bind to the maternal ribonuclease and prevent degradation of zygotic transcripts (Tong et al., *supra*). In addition to its role in oocyte development and embryogenesis, MATER may also be relevant to the pathogenesis of ovarian immunity, as it is a target of autoantibodies in mice with autoimmune oophoritis (Tong and Nelson, *supra*).

The maternal mRNA D7 is a moderately abundant transcript in *Xenopus laevis* whose expression is highest in, and perhaps restricted to, oogenesis and early embryogenesis. The D7 protein is absent from oocytes and first begins to accumulate during oocyte maturation. Its levels are highest during the first day of embryonic development and then they decrease. The loss of D7 protein affects the maturation process itself, significantly delaying the time course of germinal vesicle breakdown. Thus, D7 is a newly described protein involved in oocyte maturation (Smith, R.C. et al. (1988) Genes Dev. 2(10):1296-306.)

Many other genes are involved in subsequent stages of embryogenesis. After fertilization, the oocyte is guided by fimbria at the distal end of each fallopian tube into and through the fallopian tube and thence into the uterus. Changes in the uterine endometrium prepare the tissue to support the implantation and embryonic development of a fertilized ovum. Several stages of division have

5   occurred before the dividing ovum, now a blastocyst with about 100 cells, enters the uterus. Upon reaching the uterus, the developing blastocyst usually remains in the uterine cavity an additional two to four days before implanting in the endometrium, the inner lining of the uterus. Implantation results from the action of trophoblast cells that develop over the surface of the blastocyst. These cells secrete proteolytic enzymes that digest and liquefy the cells of the endometrium. The invasive

10 process is reviewed in Fisher, S.J. and C.H. Damsky (1993; Semin Cell Biol 4:183-188) and Graham, C.H. and P.K. Lala (1992; Biochem Cell Biol 70:867-874). Once implantation has taken place, the trophoblast and other sublying cells proliferate rapidly, forming the placenta and the various membranes of pregnancy. (See Guyton, A.C. (1991) Textbook of Medical Physiology, 8th ed. W.B. Saunders Company, Philadelphia PA, pp. 915-919.)

15   The placenta has an essential role in protecting and nourishing the developing fetus. In most species the syncytiotrophoblast layer is present on the outside of the placenta at the fetal-maternal interface. This is a continuous structure, one cell deep, formed by the fusion of the constituent trophoblast cells. The syncytiotrophoblast cells play important roles in maternal-fetal exchange, in tissue remodeling during fetal development, and in protecting the developing fetus from the maternal

20 immune response (Stoye, J.P. and J.M. Coffin (2000) Nature 403:715-717).

A gene called syncytin is the envelope gene of a human endogenous defective provirus. Syncytin is expressed in high levels in placenta, and more weakly in testis, but is not detected in any other tissues (Mi, S. et al. (2000) Nature 403:785-789). Syncytin expression in the placenta is restricted to the syncytiotrophoblasts. Since retroviral *env* proteins are often involved in promoting

25 cell fusion events, it was thought that syncytin might be involved in regulating the fusion of trophoblast cells into the syncytiotrophoblast layer. Experiments demonstrated that syncytin can mediate cell fusion *in vitro*, and that anti-syncytin antibodies can inhibit the fusion of placental cytotrophoblasts (Mi et al., *supra*). In addition, a conserved immunosuppressive domain present in retroviral envelope proteins, and found in syncytin at amino acid residues 373-397, might be involved

30 in preventing maternal immune responses against the developing embryo.

Syncytin may also be involved in regulating trophoblast invasiveness by inducing trophoblast fusion and terminal differentiation (Mi et al., *supra*). Insufficient trophoblast infiltration of the uterine wall is associated with placental disorders such as preeclampsia, or pregnancy induced hypertension, while uncontrolled trophoblast invasion is observed in choriocarcinoma and other

35 gestational trophoblastic diseases. Thus syncytin function may be involved in these diseases.

Cell Differentiation

Multicellular organisms are comprised of diverse cell types that differ dramatically both in structure and function, despite the fact that each cell is like the others in its hereditary endowment. Cell differentiation is the process by which cells come to differ in their structure and physiological function. The cells of a multicellular organism all arise from mitotic divisions of a single-celled zygote. The zygote is totipotent, meaning that it has the ability to give rise to every type of cell in the adult body. During development the cellular descendants of the zygote lose their totipotency and become determined. Once its prospective fate is achieved, a cell is said to have differentiated. All descendants of this cell will be of the same type.

Human growth and development requires the spatial and temporal regulation of cell differentiation, along with cell proliferation and regulated cell death. These processes coordinate to control reproduction, aging, embryogenesis, morphogenesis, organogenesis, and tissue repair and maintenance. The processes involved in cell differentiation are also relevant to disease states such as cancer, in which case the factors regulating normal cell differentiation have been altered, allowing the cancerous cells to proliferate in an anaplastic, or undifferentiated, state.

The mechanisms of differentiation involve cell-specific regulation of transcription and translation, so that different genes are selectively expressed at different times in different cells. Genetic experiments using the fruit fly *Drosophila melanogaster* have identified regulated cascades of transcription factors which control pattern formation during development and differentiation. These include the homeotic genes, which encode transcription factors containing homeobox motifs. The products of homeotic genes determine how the insect's imaginal discs develop from masses of undifferentiated cells to specific segments containing complex organs. These proteins are critical for specifying the anterior-posterior body axis during development. Many genes found to be involved in cell differentiation and development in Drosophila have homologs in mammals. Some human genes have equivalent developmental roles to their Drosophila homologs. The human homolog of the Drosophila eyes absent gene (eya) underlies branchio-oto-renal syndrome, a developmental disorder affecting the ears and kidneys (Abdelhak, S. et al. (1997) Nat. Genet. 15:157-164). In Drosophila, Iroquois homeobox genes play roles in the control of proneural and vein-forming genes, the positional identity of sensory neurons, alula and notum formation, and the establishment of a dorsoventral pattern organizing center necessary for follicle, head and eye development. Mammalian Iroquois homeobox genes are involved in the development of the nervous system and heart (Christoffels, V.M. et al. (2000) Dev. Biol. 224:263-274).

Tesmin is a testis-specific metallothionein-like protein that is specifically expressed in spermatocytes and may play a role in spermatogenesis (Sugihara, T. et al. (1999) Genomics 57:130-136). Metallothionein proteins are low-molecular-weight cysteine-rich proteins that bind to heavy

metal ions. Metallothionein proteins may bind and sequester toxic heavy metals, act as metal storage
molecules, and control concentrations of essential heavy metals, particularly zinc. By regulating the
availability of metal ions such as zinc required for certain enzymes and transcription factors,
metallothioneins may have roles in cell growth and differentiation.

5   At the cellular level, growth and development are governed by the cell's decision to enter
into or exit from the cell cycle and by the cell's commitment to a terminally differentiated state.
Differential gene expression within cells is triggered in response to extracellular signals and other
environmental cues. Such signals include growth factors and other mitogens such as retinoic acid;
cell-cell and cell-matrix contacts; and environmental factors such as nutritional signals, toxic
10   substances, and heat shock. Candidate genes that may play a role in differentiation can be identified
by altered expression patterns upon induction of cell differentiation *in vitro*.

The final step in cell differentiation results in a specialization that is characterized by the
production of particular proteins, such as contractile proteins in muscle cells, serum proteins in liver
cells and globins in red blood cell precursors. The expression of these specialized proteins depends at
15   least in part on cell-specific transcription factors. For example, the homeobox-containing
transcription factor PAX-6 is essential for early eye determination, specification of ocular tissues, and
normal eye development in vertebrates.

In the case of epidermal differentiation, the induction of differentiation-specific genes occurs
either together with or following growth arrest and is believed to be linked to the molecular events
20   that control irreversible growth arrest. Irreversible growth arrest is an early event which occurs when
cells transit from the basal to the innermost suprabasal layer of the skin and begin expressing
squamous-specific genes. These genes include those involved in the formation of the cross-linked
envelope, such as transglutaminase I and III, involucrin, loricin, and small proline-rich repeat (SPRR)
proteins. The SPRR proteins are 8-10 kDa in molecular mass, rich in proline, glutamine, and
25   cysteine, and contain similar repeating sequence elements. The SPRR proteins may be structural
proteins with a strong secondary structure or metal-binding proteins such as metallothioneins.
(Jetten, A.M. and B.L. Harvat (1997) J. Dermatol. 24:711-725; PRINTS Entry PR00021 PRORICH
Small proline-rich protein signature.)

The Wnt gene family of secreted signaling molecules is highly conserved throughout
30   eukaryotic cells. Members of the Wnt family are involved in regulating chondrocyte differentiation
within the cartilage template. Wnt-5a, Wnt-5b and Wnt-4 genes are expressed in chondrogenic
regions of the chicken limb, Wnt-5a being expressed in the perichondrium (mesenchymal cells
immediately surrounding the early cartilage template). Wnt-5a misexpression delays the maturation
of chondrocytes and the onset of bone collar formation in chicken limb (Hartmann, C. and C.J. Tabin
35   (2000) Development 127:3141-3159).

Glypicans are a family of cell surface heparan sulfate proteoglycans that play an important role in cellular growth control and differentiation. Cerebroglycan, a heparan sulfate proteoglycan expressed in the nervous system, is involved with the motile behavior of developing neurons (Stipp, C.S. et al. (1994) J. Cell Biol. 124:149-160).

5   Notch plays an active role in the differentiation of glial cells, and influences the length and organization of neuronal processes (for a review, see Frisen, J. and U. Lendahl (2001) Bioessays 23:3-7). The Notch receptor signaling pathway is important for morphogenesis and development of many organs and tissues in multicellular species. *Drosophila* fringe proteins modulate the activation of the Notch signal transduction pathway at the dorsal-ventral boundary of the wing imaginal disc.

10   Mammalian fringe-related family members participate in boundary determination during segmentation (Johnston, S.H. et al. (1997) Development 124:2245-2254).

Recently a number of proteins have been found to contain a conserved cysteine-rich domain of about 60 amino-acid residues called the LIM domain (for Lin-11 Isl-1 Mec-3) (Freyd, G. et al. (1990) Nature 344:876-879; Baltz, R. et al. (1992) Plant Cell 4:1465-1466). In the LIM domain, there

15   are seven conserved cysteine residues and a histidine. The LIM domain binds two zinc ions (Michelsen, J.W. et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90:4404-4408). LIM does not bind DNA; rather, it seems to act as an interface for protein-protein interaction.

Apoptosis

Apoptosis is the genetically controlled process by which unneeded or defective cells undergo

20   programmed cell death. Selective elimination of cells is as important for morphogenesis and tissue remodeling as is cell proliferation and differentiation. Lack of apoptosis may result in hyperplasia and other disorders associated with increased cell proliferation. Apoptosis is also a critical component of the immune response. Immune cells such as cytotoxic T-cells and natural killer cells prevent the spread of disease by inducing apoptosis in tumor cells and virus-infected cells. In

25   addition, immune cells that fail to distinguish self molecules from foreign molecules must be eliminated by apoptosis to avoid an autoimmune response.

Apoptotic cells undergo distinct morphological changes. Hallmarks of apoptosis include cell shrinkage, nuclear and cytoplasmic condensation, and alterations in plasma membrane topology. Biochemically, apoptotic cells are characterized by increased intracellular calcium concentration,

30   fragmentation of chromosomal DNA, and expression of novel cell surface components.

The molecular mechanisms of apoptosis are highly conserved, and many of the key protein regulators and effectors of apoptosis have been identified. Apoptosis generally proceeds in response to a signal which is transduced intracellularly and results in altered patterns of gene expression and protein activity. Signaling molecules such as hormones and cytokines are known both to stimulate

35   and to inhibit apoptosis through interactions with cell surface receptors. Transcription factors also

play an important role in the onset of apoptosis. A number of downstream effector molecules, especially proteases, have been implicated in the degradation of cellular components and the proteolytic activation of other apoptotic effectors.

5   The Bcl-2 family of proteins, as well as other cytoplasmic proteins, are key regulators of apoptosis. There are at least 15 Bcl-2 family members within 3 subfamilies. These proteins have been identified in mammalian cells and in viruses, and each possesses at least one of four Bcl-2 homology domains (BH1 to BH4), which are highly conserved. Bcl-2 family proteins contain the BH1 and BH2 domains, which are found in members of the pro-survival subfamily, while those proteins which are most similar to Bcl-2 have all four conserved domains, enabling inhibition of

10   apoptosis following encounters with a variety of cytotoxic challenges. Members of the pro-survival subfamily include Bcl-2, Bcl-x_L, Bcl-w, Mcl-1, and A1 in mammals; NF-13 (chicken); CED-9 (*Caenorhabditis elegans*); and viral proteins BHRF1, LMW5-HL, ORF16, KS-Bcl-2, and E1B-19K. The BH3 domain is essential for the function of pro-apoptosis subfamily proteins. The two pro-apoptosis subfamilies, Bax and BH3, include Bax, Bak, and Bok (also called Mtd); and Bik, Blk, Hrk,

15   BNIP3, Bim$_L$, Bad, Bid, and Egl-1 (*C. elegans*); respectively. Members of the Bax subfamily contain the BH1, BH2, and BH3 domains, and resemble Bcl-2 rather closely. In contrast, members of the BH3 subfamily have only the 9-16 residue BH3 domain, being otherwise unrelated to any known protein, and only Bik and Blk share sequence similarity. The proteins of the two pro-apoptosis subfamilies may be the antagonists of pro-survival subfamily proteins. This is illustrated in *C.*

20   *elegans* where Egl-1, which is required for apoptosis, binds to and acts via CED-9 (for review, see Adams, J.M. and S. Cory (1998) Science 281:1322-1326).

Heterodimerization between pro-apoptosis and anti-apoptosis subfamily proteins seems to have a titrating effect on the functions of these protein subfamilies, which suggests that relative concentrations of the members of each subfamily may act to regulate apoptosis. Heterodimerization

25   is not required for a pro-survival protein; however, it is essential in the BH3 subfamily, and less so in the Bax subfamily.

The Bcl-2 protein has 2 isoforms, alpha and beta, which are formed by alternative splicing. It forms homodimers and heterodimers with Bax and Bak proteins and the Bcl-X isoform Bcl-x_S. Heterodimerization with Bax requires intact BH1 and BH2 domains, and is necessary for pro-survival

30   activity. The BH4 domain seems to be involved in pro-survival activity as well. Bcl-2 is located within the inner and outer mitochondrial membranes, as well as within the nuclear envelope and endoplasmic reticulum, and is expressed in a variety of tissues. Its involvement in follicular lymphoma (type II chronic lymphatic leukemia) is seen in a chromosomal translocation T(14;18) (q32;q21) and involves immunoglobulin gene regions.

35   The Bcl-x protein is a dominant regulator of apoptotic cell death. Alternative splicing results

in three isoforms, Bcl-xB, a long isoform, and a short isoform. The long isoform exhibits cell death repressor activity, while the short isoform promotes apoptosis. Bcl-xL forms heterodimers with Bax and Bak, although heterodimerization with Bax does not seem to be necessary for pro-survival (anti-apoptosis) activity. Bcl-xS forms heterodimers with Bcl-2. Bcl-x is found in mitochondrial

5   membranes and the perinuclear envelope. Bcl-xS is expressed at high levels in developing lymphocytes and other cells undergoing a high rate of turnover. Bcl-xL is found in adult brain and in other tissues' long-lived post-mitotic cells. As with Bcl-2, the BH1, BH2, and BH4 domains are involved in pro-survival activity.

The Bcl-w protein is found within the cytoplasm of almost all myeloid cell lines and in

10   numerous tissues, with the highest levels of expression in brain, colon, and salivary gland. This protein is expressed in low levels in testis, liver, heart, stomach, skeletal muscle, and placenta, and a few lymphoid cell lines. Bcl-w contains the BH1, BH2, and BH4 domains, all of which are needed for its cell survival promotion activity. Although mice in which Bcl-w gene function was disrupted by homologous recombination were viable, healthy, and normal in appearance, and adult females had

15   normal reproductive function, the adult males were infertile. In these males, the initial, prepuberty stage of spermatogenesis was largely unaffected and the testes developed normally. However, the seminiferous tubules were disorganized, contained numerous apoptotic cells, and were incapable of producing mature sperm. This mouse model may be applicable to some cases of human male sterility and suggests that alteration of programmed cell death in the testes may be useful in modulating

20   fertility (Print, C.G. et al. (1998) Proc. Natl. Acad. Sci. USA 95:12424-12431).

Studies in rat ischemic brain found Bcl-w to be overexpressed relative to its normal low constitutive level of expression in nonischemic brain. Furthermore, in vitro studies to examine the mechanism of action of Bcl-w revealed that isolated rat brain mitochondria were unable to respond to an addition of recombinant Bax or high concentrations of calcium when Bcl-w was also present. The

25   normal response would be the release of cytochrome c from the mitochondria. Additionally, recombinant Bcl-w protein was found to inhibit calcium-induced loss of mitochondrial transmembrane potential, which is indicative of permeability transition. Together these findings suggest that Bcl-w may be a neuro-protectant against ischemic neuronal death and may achieve this protection via the mitochondrial death-regulatory pathway (Yan, C. et al. (2000) J. Cereb. Blood Flow

30   Metab. 20:620-630).

The bfl-1 gene is an additional member of the Bcl-2 family, and is also a suppressor of apoptosis. The Bfl-1 protein has 175 amino acids, and contains the BH1, BH2, and BH3 conserved domains found in Bcl-2 family members. It also contains a Gln-rich NH2-terminal region and lacks an NH domain 1, unlike other Bcl-2 family members. The mouse A1 protein shares high sequence

35   homology with Bfl-1 and has the 3 conserved domains found in Bfl-1. Apoptosis induced by the p53

tumor suppressor protein is suppressed by Bfl-1, similar to the action of Bcl-2, Bcl-xL, and EBV-BHRF1 (D'Sa-Eipper, C. et al. (1996) Cancer Res. 56:3879-3882). Bfl-1 is found intracellularly, with the highest expression in the hematopoietic compartment, i.e. blood, spleen, and bone marrow; moderate expression in lung, small intestine, and testis; and minimal expression in other tissues. It is

5   also found in vascular smooth muscle cells and hematopoietic malignancies. A correlation has been noted between the expression level of bfl-1 and the development of stomach cancer, suggesting that the Bfl-1 protein is involved in the development of stomach cancer, either in the promotion of cancerous cell survival or in cancer (Choi, S.S. et al. (1995) Oncogene 11:1693-1698).

Cancers are characterized by continuous or uncontrolled cell proliferation. Some cancers are

10   associated with suppression of normal apoptotic cell death. Strategies for treatment may involve either reestablishing control over cell cycle progression, or selectively stimulating apoptosis in cancerous cells (Nigg, E.A. (1995) BioEssays 17:471-480). Immunological defenses against cancer include induction of apoptosis in mutant cells by tumor suppressors, and the recognition of tumor antigens by T lymphocytes. Response to mitogenic stresses is frequently controlled at the level of

15   transcription and is coordinated by various transcription factors. For example, the Rel/NF-kappa B family of vertebrate transcription factors plays a pivotal role in inflammatory and immune responses to radiation. The NF-kappa B family includes p50, p52, RelA, RelB, cRel, and other DNA-binding proteins. The p52 protein induces apoptosis, upregulates the transcription factor c-Jun, and activates c-Jun N-terminal kinase 1 (JNK1) (Sun, L. et al. (1998) Gene 208:157-166). Most NF-kappa B

20   proteins form DNA-binding homodimers or heterodimers. Dimerization of many transcription factors is mediated by a conserved sequence known as the bZIP domain, characterized by a basic region followed by a leucine zipper.

The Fas/Apo-1 receptor (FAS) is a member of the tumor necrosis factor (TNF) receptor family. Upon binding its ligand (Fas ligand), the membrane-spanning FAS induces apoptosis by

25   recruiting several cytoplasmic proteins that transmit the death signal. One such protein, termed FAS-associated protein factor 1 (FAF1), was isolated from mice, and it was demonstrated that expression of FAF1 in L cells potentiated FAS-induced apoptosis (Chu, K. et al. (1995) Proc. Natl. Acad. Sci. USA 92:11894-11898). Subsequently, FAS-associated factors have been isolated from numerous other species, including fruit fly and quail (Frohlich, T. et al. (1998) J. Cell Sci. 111:2353-2363).

30   Another cytoplasmic protein that functions in the transmittal of the death signal from Fas is the Fas-associated death domain protein, also known as FADD. FADD transmits the death signal in both FAS-mediated and TNF receptor-mediated apoptotic pathways by activating caspase-8 (Bang, S. et al. (2000) J. Biol. Chem. 275:36217-36222).

Fragmentation of chromosomal DNA is one of the hallmarks of apoptosis. DNA

35   fragmentation factor (DFF) is a protein composed of two subunits, a 40-kDa caspase-activated

nuclease termed DFF40/CAD, and its 45-kDa inhibitor DFF45/ICAD. Two mouse homologs of DFF45/ICAD, termed CIDE-A and CIDE-B, have recently been described (Inohara, N. et al. (1998) EMBO J. 17:2526-2533). CIDE-A and CIDE-B expression in mammalian cells activated apoptosis, while expression of CIDE-A alone induced DNA fragmentation. In addition, FAS-mediated

5   apoptosis was enhanced by CIDE-A and CIDE-B, further implicating these proteins as effectors that mediate apoptosis.

Transcription factors play an important role in the onset of apoptosis. A number of downstream effector molecules, particularly proteases such as the cysteine proteases called caspases, are involved in the initiation and execution phases of apoptosis. The activation of the caspases results

10   from the competitive action of the pro-survival and pro-apoptosis Bcl-2-related proteins (Print, C.G. et al. (1998) Proc. Natl. Acad. Sci. USA 95:12424-12431). A pro-apoptotic signal can activate initiator caspases that trigger a proteolytic caspase cascade, leading to the hydrolysis of target proteins and the classic apoptotic death of the cell. Two active site residues, a cysteine and a histidine, have been implicated in the catalytic mechanism. Caspases are among the most specific

15   endopeptidases, cleaving after aspartate residues.

Caspases are synthesized as inactive zymogens consisting of one large (p20) and one small (p10) subunit separated by a small spacer region, and a variable N-terminal prodomain. This prodomain interacts with cofactors that can positively or negatively affect apoptosis. An activating signal causes autoproteolytic cleavage of a specific aspartate residue (D297 in the caspase-1

20   numbering convention) and removal of the spacer and prodomain, leaving a p10/p20 heterodimer. Two of these heterodimers interact via their small subunits to form the catalytically active tetramer. The long prodomains of some caspase family members have been shown to promote dimerization and auto-processing of procaspases. Some caspases contain a "death effector domain" in their prodomain by which they can be recruited into self-activating complexes with other caspases and FADD protein-

25   associated death receptors or the TNF receptor complex. In addition, two dimers from different caspase family members can associate, changing the substrate specificity of the resultant tetramer.

Tumor necrosis factor (TNF) and related cytokines induce apoptosis in lymphoid cells. (Reviewed in Nagata, S. (1997) Cell 88:355-365.) Binding of TNF to its receptor triggers a signal transduction pathway that results in the activation of a proteolytic caspase cascade. One such

30   caspase, ICE (Interleukin-1β converting enzyme), is a cysteine protease comprised of two large and two small subunits generated by ICE auto-cleavage (Dinarello, C.A. (1994) FASEB J. 8:1314-1325). ICE is expressed primarily in monocytes. ICE processes the cytokine precursor, interleukin-1β, into its active form, which plays a central role in acute and chronic inflammation, bone resorption, myelogenous leukemia, and other pathological processes. ICE and related caspases cause apoptosis

35   when overexpressed in transfected cell lines.

A caspase recruitment domain (CARD) is found within the prodomain of several apical caspases and is conserved in several apoptosis regulatory molecules such as Apaf-2, RAIDD, and cellular inhibitors of apoptosis proteins (IAPs) (Hofmann, K. et al. (1997) Trends Biochem. Sci. 22:155-157). The regulatory role of CARD in apoptosis may be to allow proteins such as Apaf-1 to

5   associate with caspase-9 (Li, P. et al. (1997) Cell 91:479-489). A human cDNA encoding an apoptosis repressor with a CARD (ARC) which is expressed in both skeletal and cardiac muscle has been identified and characterized. ARC functions as an inhibitor of apoptosis and interacts selectively with caspases (Koseki, T. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5156-5160). All of these interactions have clear effects on the control of apoptosis (reviewed in Chan S.L. and M.P.

10   Mattson (1999) J. Neurosci. Res. 58:167-190; Salveson, G.S. and V.M. Dixit (1999) Proc. Natl. Acad. Sci. USA 96:10964-10967).

ES18 was identified as a potential regulator of apoptosis in mouse T-cells (Park, E.J. et al. (1999) Nuc. Acid. Res. 27:1524-1530). ES18 is 428 amino acids in length, contains an N-terminal proline-rich region, an acidic glutamic acid-rich domain, and a putative LXXLL nuclear receptor

15   binding motif. The protein is preferentially expressed in lymph nodes and thymus. The level of ES18 expression increases in T-cell thymoma S49.1 in response to treatment with dexamethasone, staurosporine, or C2-ceramide, which induce apoptosis. ES18 may play a role in stimulating apoptotic cell death in T-cells.

The rat ventral prostate (RVP) is a model system for the study of hormone-regulated

20   apoptosis. RVP epithelial cells undergo apoptosis in response to androgen deprivation. Messenger RNA (mRNA) transcripts that are up-regulated in the apoptotic RVP have been identified (Briehl, M. M. and R.L. Miesfeld (1991) Mol. Endocrinol. 5:1381-1388). One such transcript encodes RVP.1, the precise role of which in apoptosis has not been determined. The human homolog of RVP.1, hRVP1, is 89% identical to the rat protein (Katahira, J. et al. (1997) J. Biol. Chem. 272:26652-

25   26658). hRVP1 is 220 amino acids in length and contains four transmembrane domains. hRVP1 is highly expressed in the lung, intestine, and liver. Interestingly, hRVP1 functions as a low affinity receptor for the *Clostridium perfringens* enterotoxin, a causative agent of diarrhea in humans and other animals.

Cytokine-mediated apoptosis plays an important role in hematopoiesis and the immune

30   response. Myeloid cells, which are the stem cell progenitors of macrophages, neutrophils, erythrocytes, and other blood cells, proliferate in response to specific cytokines such as granulocyte/macrophage-colony stimulating factor (GM-CSF) and interleukin-3 (IL-3). When deprived of GM-CSF or IL-3, myeloid cells undergo apoptosis. The murine *requiem* (*req*) gene encodes a putative transcription factor required for this apoptotic response in the myeloid cell line

35   FDCP-1 (Gabig, T. G. et al. (1994) J. Biol. Chem. 269:29515-29519). The Req protein is 371 amino

acids in length and contains a nuclear localization signal, a single *Kruppel*-type zinc finger, an acidic domain, and a cluster of four unique zinc-finger motifs enriched in cysteine and histidine residues involved in metal binding. Expression of *req* is not myeloid- or apoptosis-specific, suggesting that additional factors regulate Req activity in myeloid cell apoptosis.

5   Dysregulation of apoptosis has recently been recognized as a significant factor in the pathogenesis of many human diseases. For example, excessive cell survival caused by decreased apoptosis can contribute to disorders related to cell proliferation and the immune response. Such disorders include cancer, autoimmune diseases, viral infections, and inflammation. In contrast, excessive cell death caused by increased apoptosis can lead to degenerative and immunodeficiency

10   disorders such as AIDS, neurodegenerative diseases, and myelodysplastic syndromes. (Thompson, C.B. (1995) Science 267:1456-1462.)

Impaired regulation of apoptosis is also associated with loss of neurons in Alzheimer's disease. Alzheimer's disease is a progressive neurodegenerative disorder that is characterized by the formation of senile plaques and neurofibrillary tangles containing amyloid beta peptide. These

15   plaques are found in limbic and association cortices of the brain, including hippocampus, temporal cortices, cingulate cortex, amygdala, nucleus basalis and locus caeruleus. B-amyloid peptide participates in signaling pathways that induce apoptosis and lead to the death of neurons (Kajkowski, C. et al. (2001) J. Biol. Chem. 276:18748-18756). Early in Alzheimer's pathology, physiological changes are visible in the cingulate cortex (Minoshima, S. et al. (1997) Annals of Neurology 42:85-

20   94). In subjects with advanced Alzheimer's disease, accumulating plaques damage the neuronal architecture in limbic areas and eventually cripple the memory process.

Cell Differentiation and Proliferation

Tissue growth involves complex and ordered patterns of cell proliferation, cell differentiation, and apoptosis. Cell proliferation must be regulated to maintain both the number of

25   cells and their spatial organization. This regulation depends upon the appropriate expression of proteins which control cell cycle progression in response to extracellular signals, such as growth factors and other mitogens, and intracellular cues, such as DNA damage or nutrient starvation. Molecules which directly or indirectly modulate cell cycle progression fall into several categories, including growth factors and their receptors, second messenger and signal transduction proteins,

30   oncogene products, tumor-suppressor proteins, and mitosis-promoting factors.

Growth factors were originally described as serum factors required to promote cell proliferation. Most growth factors are large, secreted polypeptides that act on cells in their local environment. Growth factors bind to and activate specific cell surface receptors and initiate intracellular signal transduction cascades. Many growth factor receptors are classified as receptor

35   tyrosine kinases which undergo autophosphorylation upon ligand binding. Autophosphorylation

enables the receptor to interact with signal transduction proteins characterized by the presence of SH2 or SH3 domains (Src homology regions 2 or 3). These proteins then modulate the activity state of small G-proteins, such as Ras, Rab, and Rho, along with GTPase activating proteins (GAPs), guanine nucleotide releasing proteins (GNRPs), and other guanine nucleotide exchange factors. Small G

5   proteins act as molecular switches that activate other downstream events, such as mitogen-activated protein kinase (MAP kinase) cascades. MAP kinases ultimately activate transcription of mitosis-promoting genes.

In addition to growth factors, small signaling peptides and hormones also influence cell proliferation. These molecules bind primarily to another class of receptor, the trimeric G-protein

10   coupled receptor (GPCR), found predominantly on the surface of immune, neuronal and neuroendocrine cells. Upon ligand binding, the GPCR activates a trimeric G protein which in turn triggers increased levels of intracellular second messengers such as phospholipase C, Ca2+, and cyclic AMP. Most GPCR-mediated signaling pathways indirectly promote cell proliferation by causing the secretion or breakdown of other signaling molecules that have direct mitogenic effects.

15   These signaling cascades often involve activation of kinases and phosphatases. Some growth factors, such as some members of the transforming growth factor beta (TGF-β) family, act on some cells to stimulate cell proliferation and on other cells to inhibit it. Growth factors may also stimulate a cell at one concentration and inhibit the same cell at another concentration. Most growth factors also have a multitude of other actions besides the regulation of cell growth and division: they can control the

20   proliferation, survival, differentiation, migration, or function of cells depending on the circumstance. For example, the tumor necrosis factor/nerve growth factor (TNF/NGF) family can activate or inhibit cell death, as well as regulate proliferation and differentiation. The cell response depends on the type of cell, its stage of differentiation and transformation status, which surface receptors are stimulated, and the types of stimuli acting on the cell (Smith, A. et al. (1994) Cell 76:959-962; and Nocentini, G.

25   et al. (1997) Proc. Natl. Acad. Sci. USA 94:6216-6221).

Neighboring cells in a tissue compete for growth factors, and when provided with "unlimited" quantities in a perfused system will grow to even higher cell densities before reaching density-dependent inhibition of cell division. Cells often demonstrate an anchorage dependence of cell division as well. This anchorage dependence may be associated with the formation of focal contacts

30   linking the cytoskeleton with the extracellular matrix (ECM). The expression of ECM components can be stimulated by growth factors. For example, TGF-β stimulates fibroblasts to produce a variety of ECM proteins, including fibronectin, collagen, and tenascin (Pearson, C.A. et al. (1988) EMBO J. 7:2677-2981). In fact, for some cell types specific ECM molecules, such as laminin or fibronectin, may act as growth factors. Tenascin-C and -R, expressed in developing and lesioned neural tissue,

35   provide stimulatory/anti-adhesive or inhibitory properties, respectively, for axonal growth (Faissner,

A. (1997) Cell Tissue Res. 290:331-341).

Oncogenes

Cancers are associated with the activation of oncogenes which are derived from normal cellular genes. These oncogenes encode oncoproteins which convert normal cells into malignant

5   cells. Some oncoproteins are mutant isoforms of the normal protein, and other oncoproteins are abnormally expressed with respect to location or amount of expression. The latter category of oncoprotein causes cancer by altering transcriptional control of cell proliferation. Five classes of oncoproteins are known to affect cell cycle controls. These classes include growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control

10   proteins. Viral oncogenes are integrated into the human genome after infection of human cells by certain viruses. Examples of viral oncogenes include v-src, v-abl, and v-fps. Many cases related to the overexpression of proteins associated with tumors and metastasis have been reported. The Mta1 gene has been cloned in mice, in both cell lines and tissues representing metastatic tumors (Simpson, A. et al. (2001) Gene 273:29-39). Expression of the melanoma antigen-encoding gene (MAGE)

15   family of proteins has also been detected in many tumors. GAC1, a new member of the leucine-rich repeat superfamily, is amplified and overexpressed in malignant gliomas (Almeida, A. et al. (1998) Oncogene 16:2997-3002).

Many oncogenes have been identified and characterized. These include sis, erbA, erbB, her-2, mutated G_s, src, abl, ras, crk, jun, fos, myc, and mutated tumor-suppressor genes such as RB, p53,

20   mdm2, Cip1, p16, and cyclin D. Transformation of normal genes to oncogenes may also occur by chromosomal translocation. The Philadelphia chromosome, characteristic of chronic myeloid leukemia and a subset of acute lymphoblastic leukemias, results from a reciprocal translocation between chromosomes 9 and 22 that moves a truncated portion of the proto-oncogene c-abl to the breakpoint cluster region (bcr) on chromosome 22.

25   Tumor-suppressor genes are involved in regulating cell proliferation. Mutations which cause reduced or loss of function in tumor-suppressor genes result in uncontrolled cell proliferation. For example, the retinoblastoma gene product (RB), in a non-phosphorylated state, binds several early-response genes and suppresses their transcription, thus blocking cell division. Phosphorylation of RB causes it to dissociate from the genes, releasing the suppression, and allowing cell division to

30   proceed.

Angiogenesis

Angiogenesis is the process by which new capillaries are formed by sprouting from preexisting vessels. It is a vital function for the gowth of normal tissues during embryogenesis as well as for the pathological growth of tumors. Pathological proliferation fo cancer cells will not

35   result in a proportional increse in mass without access to the blood circulation. Tumors form their

own circulatory system by upregulating angiogenic stimulators and by downregulation angiogenesis inhibitors. The inability of metastases to induce an angiogenic response results in a dormant phenotype. Angiostatin is a circulating inhibitor of angiogenesis. In vitro, it inhibits endothelial cell migration, proliferation, and tube formation, and induces apoptosis in a cell type-specific manner.

5   Angiomotin is an angiostatin-binding peptide that mediates angiostatin inhibition of migration and tube formation of endothelial cells. Angiomotin is expressed in the endothelial cells of capillaries and the larger vessels of the human placenta. Angiostatin inhibits cell migration by interfering with angiomotin activity in endothelial cells (Troyanovsky, B. et al. (2001) J. Cell Biol. 152:1247-1254).

Nucleolar protein p120 is a proliferation-associated antigen expressed by cells in early G1

10   phase, identified by the monoclonal antibody FB. It is very cancer specific. In fact, quantitative immunohistochemical analysis of p120 protein is an easy and reliable method for the assessment of clinical outcome and the definition of risk groups in oral carcinoma (Ventura, L. et al. (1999) Anticancer Res. 19:1423-1426). P120 contains a basic domain, an acidic domain, a hydrophobic and methionine-rich domain, and a domain rich in cysteine and proline residues (Fonagy, A. et al. (1989)

15   Cancer Commun. 1:243-251). This protein is expressed in early G1 and has not been detected in benign tumors and most normal resting tissues. Sato et al. show that the expression level of p120 in tumor tissues can be used as an independent and powerful prognostic marker for resected lung adenocarcinoma (Sato, G. et al. (1999) J. Clin. Oncol. 17:2721-2727).

The human LGI1 gene is a leucine-rich, repeat-containing gene that was cloned from the

20   t(10:19) breakpoint of the T98G glioblastoma cell line. The LGI1 gene maps to 10q24, a region of peak LOH in malignant gliomas, and is inactivated during the transition from low to high-grade brain tumors. The mouse lgil gene is 97% homologous to the human gene at the amino acid level and 91% homologous at the nucleotide level. LGI1 contains 8 exons, where each of the four leucine-rich repeat units is contained in an individual 72-bp exon. The cysteine-rich regions flanking the LRR and

25   the single trans-membrane domain do not occupy individual exons (Somerville, R.P. et al. (2000) Mamm. Genome 11:622-627).

Viral oncogenes are integrated into the human genome after infection of human cells by certain viruses. Examples of viral oncogenes include v-src, v-abl, and v-fps. Transformation of normal genes to oncogenes may also occur by chromosomal translocation. The Philadelphia

30   chromosome, characteristic of chronic myeloid leukemia and a subset of acute lymphoblastic leukemias, results from a reciprocal translocation between chromosomes 9 and 22 that moves a truncated portion of the proto-oncogene c-abl to the breakpoint cluster region (bcr) on chromosome 22. The hybrid c-abl-bcr gene encodes a chimeric protein that has tyrosine kinase activity. In chronic myeloid leukemia, the chimeric protein has a molecular weight of 210 kd, whereas in acute leukemias

35   a more active 180 kd tyrosine kinase is formed (Robbins, S.L. et al. (1994) Pathologic Basis of

Disease, W.B. Saunders Co., Philadelphia PA).

The Ras superfamily of small GTPases is involved in the regulation of a wide range of cellular signaling pathways. Ras family proteins are membrane-associated proteins acting as molecular switches that bind GTP and GDP, hydrolyzing GTP to GDP. In the active GTP-bound
5   state Ras family proteins interact with a variety of cellular targets to activate downstream signaling pathways. For example, members of the Ras subfamily are essential in transducing signals from receptor tyrosine kinases (RTKs) to a series of serine/threonine kinases which control cell growth and differentiation. Activated Ras genes were initially found in human cancers and subsequent studies confirmed that Ras function is critical in the determination of whether cells continue to grow or
10 become terminally differentiated (Barbacid, M. (1987) Annu. Rev. Biochem. 56:779-827; Treisman, R. (1994) Curr. Opin. Genet. Dev. 4:96-98). Mutant Ras proteins, which bind but can not hydrolyze GTP, are permanently activated, and cause continuous cell proliferation or cancer.

Activation of Ras family proteins is catalyzed by guanine nucleotide exchange factors (GEFs) which catalyze the dissociation of bound GDP and subsequent binding of GTP. A recently
15 discovered RalGEF-like protein, RGL3, interacts with both Ras and the related protein Rit. Constitutively active Rit, like Ras, can induce oncogenic transformation, although since Rit fails to interact with most known Ras effector proteins, novel cellular targets may be involved in Rit transforming activity. RGL3 interacts with both Ras and Rit, and thus may act as a downstream effector for these proteins (Shao, H. and D.A. Andres (2000) J. Biol. Chem. 275:26914-26924).

20 Tumor antigens

Tumor antigens are cell surface molecules that are differentially expressed in tumor cells relative to non-tumor tissues. Tumor antigens make tumor cells immunologically distinct from normal cells and are potential diagnostics for human cancers. Several monoclonal antibodies have been identified which react specifically with cancerous cells such as T-cell acute lymphoblastic
25 leukemia and neuroblastoma (Minegishi, M. et al. (1989) Leukemia Res. 13:43-51; Takagi, S. et al. (1995) Int. J. Cancer 61:706-715). In addition, the discovery of high level expression of the HER2 gene in breast tumors has led to the development of therapeutic treatments (Liu, E. et al. (1992) Oncogene 7: 1027-1032; Kern, J.A. (1993) Am. J. Respir. Cell Mol. Biol. 9:448-454). Tumor antigens are found on the cell surface and have been characterized either as membrane proteins or
30 glycoproteins. For example, MAGE genes encode a family of tumor antigens recognized on melanoma cell surfaces by autologous cytolytic T lymphocytes. Among the 12 human MAGE genes isolated, half are differentially expressed in tumors of various histological types (De Plaen, E. et al. (1994) Immunogenetics 40:360-369). None of the 12 MAGE genes, however, is expressed in healthy tissues except testis and placenta.

Tumor suppressors

Tumor suppressor genes are generally defined as genetic elements whose loss or inactivation contributes to the deregulation of cell proliferation and the pathogenesis and progression of cancer. Tumor suppressor genes normally function to control or inhibit cell growth in response to stress and

5   to limit the proliferative life span of the cell. Several tumor suppressor genes have been identified including the genes encoding the retinoblastoma (Rb) protein, p53, and the breast cancer 1 and 2 proteins (BRCA1 and BRCA2). Mutations in these genes are associated with acquired and inherited genetic predisposition to the development of certain cancers.

The role of p53 in the pathogenesis of cancer has been extensively studied. (Reviewed in

10 Aggarwal, M. L. et al. (1998) J. Biol. Chem. 273:1-4; Levine, A. (1997) Cell 88:323-331.) About 50% of all human cancers contain mutations in the *p53* gene. These mutations result in either the absence of functional p53 or, more commonly, a defective form of p53 which is overexpressed. p53 is a transcription factor that contains a central core domain required for DNA binding. Most cancer-associated mutations in p53 localize to this domain. In normal proliferating cells, p53 is expressed at

15 low levels and is rapidly degraded. p53 expression and activity is induced in response to DNA damage, abortive mitosis, and other stressful stimuli. In these instances, p53 induces apoptosis or arrests cell growth until the stress is removed. Downstream effectors of p53 activity include apoptosis-specific proteins and cell cycle regulatory proteins, including Rb, oncogene products, cyclins, and cell cycle-dependent kinases.

20 A novel candidate tumor suppressor gene, LGI1 (leucine-rich gene-glioma inactivated), has been identified in glioblastoma multiforme tumors using positional cloning techniques (Chernova, O.B. et al. (1998) 17:2873-2881). Glioblastoma multiforme is characterized by abundant angiogenic activity and loss of tumor suppressor genes on chromosome 10. The human LGI1 gene is a leucine-rich, repeat-containing gene cloned from the t(10;19) breakpoint of the T98G glioblastoma cell line

25 (Somerville, R.P.T., et al. (2000) Mammalian Genome 11:622-627). Specifically, LGI1 has been localized to the 10q24 region, and its association with glial tumors may be the result of a t(10;19) (q24;q13) translocation event. The LGI1 gene encodes a protein with molecular mass of 60kD and contains 3.5 leucine-rich repeats (LRR).

The metastasis-suppressor gene KAI1 (CD82) has been reported to be related to the tumor

30 suppressor gene p53. KAI1 is involved in the progression of human prostatic cancer and possibly lung and breast cancers when expression is decreased. KAI1 encodes a member of a structurally distinct family of leukocyte surface glycoproteins. The family is known as either the tetraspan transmembrane protein family or transmembrane 4 superfamily (TM4SF) as the members of this family span the plasma membrane four times. The family is composed of integral membrane proteins

having a N-terminal membrane-anchoring domain which functions as both a membrane anchor and a translocation signal during protein biosynthesis. The N-terminal membrane-anchoring domain is not cleaved during biosynthesis. TM4SF proteins have three additional transmembrane regions, seven or more conserved cysteine residues, are similar in size (218 to 284 residues), and all have a large

5   extracellular hydrophilic domain with three potential N-glycosylation sites. The promoter region contains many putative binding motifs for various transcription factors, including five AP2 sites and nine SpI sites. Gene structure comparisons of KAI1 and seven other members of the TM4SF indicate that the splicing sites relative to the different structural domains of the predicted proteins are conserved. This suggests that these genes are related evolutionarily and arose through gene

10   duplication and divergent evolution (Levy, S. et al. (1991) J. Biol. Chem. 266:14597-14602; Dong, J.T. et al. (1995) Science 268:884-886; Dong, J.T. et al., (1997) Genomics 41:25-32).

The Leucine-rich gene-Glioma Inactivated (LGI1) protein shares homology with a number of transmembrane and extracellular proteins which function as receptors and adhesion proteins. LGI1 is encoded by an LLR (leucine-rich, repeat-containing) gene and maps to 10q24. LGI1 has four LLRs

15   which are flanked by cysteine-rich regions and one transmembrane domain (Somerville, R.P. et al. (2000) Mamm. Genome 11:622-627). LGI1 expression is seen predominantly in neural tissues, especially brain. The loss of tumor suppressor activity is seen in the inactivation of the LGI1 protein which occurs during the transition from low to high-grade tumors in malignant gliomas. The reduction of LGI1 expression in low grade brain tumors and its significant reduction or absence of

20   expression in malignant gliomas suggests that it could be used for diagnosis of glial tumor progression (Chernova, O.B. et al. (1998) Oncogene 17:2873-2881).

The ST13 tumor suppressor was identified in a screen for factors related to colorectal carcinomas by subtractive hybridization between cDNA of normal mucosal tissues and mRNA of colorectal carcinoma tissues (Cao, J. et al. (1997) J. Cancer Res. Clin. Oncol. 123:447-451). ST13 is

25   down-regulated in human colorectal carcinomas.

Mutations in the von Hippel-Lindau (VHL) tumor suppressor gene are associated with retinal and central nervous system hemangioblastomas, clear cell renal carcinomas, and pheochromocytomas (Hoffman, M. et al. (2001) Hum. Mol. Genet. 10:1019-1027; Kamada, M. (2001) Cancer Res. 61:4184-4189). Tumor progression is linked to defects or inactivation of the VHL gene. VHL

30   regulates the expression of transforming growth factor-α, the GLUT-1 glucose transporter and vascular endothelial growth factor. The VHL protein associates with elongin B, elongin C, Cul2 and Rbx1 to form a complex that regulates the transcriptional activator hypoxia-inducible factor (HIF). HIF induces genes involved in angiogenesis such as vascular endothelial growth factor and platelet-derived growth factor B. Loss of control of HIF caused by defects in VHL results in the excessive

production of angiogenic peptides. VHL may play roles in inhibition of angiogenesis, cell cycle control, fibronectin matrix assembly, cell adhesion, and proteolysis.

Mutations in tumor suppressor genes are a common feature of many cancers and often appear to affect a critical step in the pathogenesis and progression of tumors. Accordingly, Chang, F. et al.

5 (1995; J. Clin. Oncol. 13:1009-1022) suggest that it may be possible to use either the gene or an antibody to the expressed protein 1) to screen patients at increased risk for cancer, 2) to aid in diagnosis made by traditional methods, and 3) to assess the prognosis of individual cancer patients. In addition, Hamada, K. et al. (1996; Cancer Res. 56:3047-3054) are investigating the introduction of p53 into cervical cancer cells via an adenoviral vector as an experimental therapy for cervical cancer.

10 The PR-domain genes were recently recognized as playing a role in human tumorigenesis. PR-domain genes normally produce two protein products: the PR-plus product, which contains the PR domain, and the PR-minus product which lacks this domain. In cancer cells, PR-plus is disrupted or overexpressed, while PR-minus is present or overexpressed. The imbalance in the amount of these two proteins appears to be an important cause of malignancy (Jiang, G.L. and S. Huang (2000) Histol.

15 Histopathol. 15:109-117).

Many neoplastic disorders in humans can be attributed to inappropriate gene transcription. Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes (Cleary, M.L. (1992) Cancer Surv. 15:89-104). Chromosomal translocations may also produce chimeric loci which fuse the coding sequence of one

20 gene with the regulatory regions of a second unrelated gene. An important class of transcriptional regulators are the zinc finger proteins. The zinc finger motif, which binds zinc ions, generally contains tandem repeats of about 30 amino acids consisting of periodically spaced cysteine and histidine residues. Examples of this sequence pattern include the C2H2-type, C4-type, and C3HC4-type zinc fingers, and the PHD domain (Lewin, B. (1990) Genes IV, Oxford University Press, New

25 York, NY, and Cell Press, Cambridge, MA, pp. 554-570; Aasland, R., et al. (1995) Trends Biochem. Sci. 20:56-59). One clinically relevant zinc-finger protein is WT1, a tumor-suppressor protein that is inactivated in children with Wilm's tumor. The oncogene bcl-6, which plays an important role in large-cell lymphoma, is also a zinc-finger protein (Papavassiliou, A.G. (1995) N. Engl. J. Med. 332:45-47).

30 Tumor responsive proteins

Cancers, also called neoplasias, are characterized by continuous and uncontrolled cell proliferation. They can be divided into three categories: carcinomas, sarcomas, and leukemias. Carcinomas are malignant growths of soft epithelial cells that may infiltrate surrounding tissues and give rise to metastatic tumors. Sarcomas may be of epithelial origin or arise from connective tissue.

Leukemias are progressive malignancies of blood-forming tissue characterized by proliferation of leukocytes and their precursors, and may be classified as myelogenous (granulocyte- or monocyte-derived) or lymphocytic (lymphocyte-derived). Tumorigenesis refers to the progression of a tumor's growth from its inception. Malignant cells may be quite similar to normal cells within the tissue of

5   origin or may be undifferentiated (anaplastic). Tumor cells may possess few nuclei or one large polymorphic nucleus. Anaplastic cells may grow in a disorganized mass that is poorly vascularized and as a result contains large areas of ischemic necrosis. Differentiated neoplastic cells may secrete the same proteins as the tissue of origin. Cancers grow, infiltrate, invade, and destroy the surrounding tissue through direct seeding of body cavities or surfaces, through lymphatic spread, or

10   through hematogenous spread. Cancer remains a major public health concern and current preventative measures and treatments do not match the needs of most patients. Understanding of the neoplastic process of tumorigenesis can be aided by the identification of molecular markers of prognostic and diagnostic importance.

Current forms of cancer treatment include the use of immunosuppressive drugs (Morisaki, T.

15   et al. (2000) Anticancer Res. 20:3363-3373; Geoerger, B. et al. (2001) Cancer Res. 61:1527-1532). The identification of proteins involved in cell signaling, and specifically proteins that act as receptors for immunosuppressant drugs, may facilitate the development of anti-tumor agents. For example, immunophilins are a family of conserved proteins found in both prokaryotes and eukaryotes that bind to immunosuppressive drugs with varying degrees of specificity. One such group of immunophilic

20   proteins is the peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) family (PPIase, rotamase). These enzymes, first isolated from porcine kidney cortex, accelerate protein folding by catalyzing the cis-trans isomerization of proline imidic peptide bonds in oligopeptides (Fischer, G. and F.X. Schmid (1990) Biochemistry 29:2205-2212). Included within the immunophilin family are the cyclophilins (e.g., peptidyl-prolyl isomerase A or PPIA) and FK-binding protein (e.g., FKBP) subfamilies.

25   Cyclophilins are multifunctional receptor proteins which participate in signal transduction activities, including those mediated by cyclosporin (or cyclosporine). The PPIase domain of each family is highly conserved between species. Although structurally distinct, these multifunctional receptor proteins are involved in numerous signal transduction pathways, and have been implicated in folding and trafficking events.

30   The immunophilin protein cyclophilin binds to the immunosuppressant drug cyclosporin A. FKBP, another immunophilin, binds to FK506 (or rapamycin). Rapamycin is an immunosuppressant agent that arrests cells in the G_1 phase of growth, inducing apoptosis. Like cyclophilin, this macrolide antibiotic (produced by *Streptomyces tsukubaensis*) acts by binding to ubiquitous, predominantly cytosolic immunophilin receptors. These immunophilin/immunosuppressant complexes (e.g.,

cyclophilin A/cyclosporin A (CypA/CsA) and FKBP12/FK506) achieve their therapeutic results through inhibition of the phosphatase calcineurin, a calcium/calmodulin-dependent protein kinase that participates in T-cell activation (Hamilton, G.S. and J.P. Steiner (1998) J. Med. Chem. 41: 5119-5143). The murine fkbp51 gene is abundantly expressed in immunological tissues, including the

5   thymus and T lymphocytes (Baughman, G. et al. (1995) Molec. Cell. Biol. 15: 4395-4402). FKBP12/rapamycin-directed immunosuppression occurs through binding to TOR (yeast) or FRAP (FKBP12-rapamycin-associated protein, in mammalian cells), the kinase target of rapamycin essential for maintaining normal cellular growth patterns. Dysfunctional TOR signaling has been linked to various human disorders including cancer (Metcalfe, S.M. et al. (1997) Oncogene 15:1635-1642;

10   Emami, S. et al. (2001) FASEB J. 15:351-361), and autoimmunity (Damoiseaux, J.G. et al. (1996) Transplantation 62:994-1001).

Several cyclophilin isozymes have been identified, including cyclophilin B, cyclophilin C, mitochondrial matrix cyclophilin, bacterial cytosolic and periplasmic PPIases, and natural-killer cell cyclophilin-related protein possessing a cyclophilin-type PPIase domain, a putative tumor-recognition

15   complex involved in the function of natural killer (NK) cells. These cells participate in the innate cellular immune response by lysing virally-infected cells or transformed cells. NK cells specifically target cells that have lost their expression of major histocompatibility complex (MHC) class I genes (common during tumorigenesis), endowing them with the potential for attenuating tumor growth. A 150-kDa molecule has been identified on the surface of human NK cells that possesses a domain

20   which is highly homologous to cyclophilin/peptidyl-prolyl cis-trans isomerase. This cyclophilin-type protein may be a component of a putative tumor-recognition complex, a NK tumor recognition sequence (NK-TR) (Anderson, S.K. et al. (1993) Proc. Natl. Acad. Sci. USA 90:542-546). The NKTR tumor recognition sequence mediates recognition between tumor cells and large granular lymphocytes (LGLs), a subpopulation of white blood cells (comprised of activated cytotoxic T cells

25   and natural killer cells) capable of destroying tumor targets. The protein product of the NKTR gene presents on the surface of LGLs and facilitates binding to tumor targets. More recently, a mouse Nktr gene and promoter region have been located on chromosome 9. The gene encodes a NK-cell-specific 150-kDa protein (NK-TR) that is homologous to cyclophilin and other tumor-responsive proteins (Simons-Evelyn, M. et al. (1997) Genomics 40:94-100).

30   Other proteins that interact with tumorigenic tissue include cytokines such as tumor necrosis factor (TNF). The TNF family of cytokines are produced by lymphocytes and macrophages, and can cause the lysis of transformed (tumor) endothelial cells. Endothelial protein 1 (Edp1) has been identified as a human gene activated transcriptionally by TNF-alpha in endothelial cells, and a TNF-alpha inducible Edp1 gene has been identified in the mouse (Swift, S. et al. (1998) Biochim. Biophys.

Acta 1442:394-398).

Crystallins are stable, long-lived structural proteins of the vertebrate eye lens. Their structure and interactions are responsible for lens transparency. Gradual accumulation of small changes to the proteins causes age-related cataracts. The lens is formed from two protein superfamilies, the alpha-

5 and beta gamma-crystallins. The proteins of both superfamilies have a basic 2-beta-sheet domain fold, with the beta gamma-domain being made from two intercalating Greek keys. The three types of crystallin proteins: alpha, beta and gamma, are each composed of a variety of subunits. (See Slingsby, C. and Clout, N.J.(1999) Eye 13:395-402; Carver, J.A. (1999) Prog. Retin. Eye Res. 18:431-462; and Graw, J. (1997) Biol. Chem. 378:1331-1348.) Some crystallins have been

10 discovered outside the eye. AIM1, a non-lens member of the betagamma-crystallin superfamily, is associated with the control of tumorigenicity in human malignant melanoma. AIM1 is a mammalian member of the betagamma superfamily with a primarily non-lens role. AIM1 is a candidate for the putative suppressor of malignant melanoma on chromosome 6, possibly exerting its effects through interactions with the cytoskeleton (Ray, M.E. et al. (1997) Proc. Natl. Acad. Sci. U S A

15 94:3229-3234).

Spermatogenesis associated factor (SPAF) is an AAA-protein (ATPase associated with diverse activities) specific to early spermatogenesis and malignant conversion. SPAF is expressed in spermatogonia and early spermatocytes in the basal compartment of the seminiferous tubules (Liu, Y. et al. (2000) Oncogene 19:1579-1588).

20 DICE1 is a tumor suppressor locus on human chromosome 13q14, that encodes a 100 kD protein similar to the carboxy-terminal half of the mouse EGF repeat transmembrane protein DBI-1. The DICE1 protein is expressed in a wide variety of fetal and adult tissues. The DBI-1 protein interferes with the mitogenic response to insulin-like growth factor 1 (IGF-I) and is implicated in anchorage-dependent growth. Expression of the DICE1 protein is reduced or undetectable in many

25 non-small cell lung carcinomas analysed (Wieland, I. (1999) Oncogene 18:4530-4537).

Disintegrins are proteins which inhibit fibrinogen interaction with blood platelet receptors expressed on the glycoprotein IIb-IIIa complex. They act by binding to the integrin glycoprotein IIb-IIIa receptor on the platelet surface and inhibit aggregation induced by ADP, thrombin, platelet-activating factor and collagen. Disintegrins are peptides of about 70 amino acid residues that contain

30 many cysteines involved in disulfide bonds (Williams, J. et al. (1990) Biochim. Biophys. Acta 1039:81-89; Dennis, M.S. et al. (1990) Proc. Natl. Acad. Sci. U.S.A. 87:2471-2475).

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support.

Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

For example, microarray analysis can be used to determine relative expression levels of polynucleotides in normal lymphocytes compared with leukemia cells. Jurkat cells are an acute T-cell leukemia cell line that have been extensively used to study gene expression in human leukemia. Treatment of Jurkat cells with PMA (a broad activator of protein kinase C-dependent pathways) and ionomycin (a calcium ionophore that permits the entry of calcium in the cell) activates two of the major signaling pathways used by mammalian cells to interact with their environment and mimics the type of secondary signaling events elicited during optimal B-cell activation. Monitoring the expression level of polynucleotides before and after the treatment of Jurkat cells with PMA and ionomycin is useful for identifying cDNAs involved in T-cell activation.

Lung cancer

Lung cancer is the leading cause of cancer death in the United States, affecting more than 100,000 men and 50,000 women each year. Nearly 90% of the patients diagnosed with lung cancer are cigarette smokers. Tobacco smoke contains thousands of noxious substances that induce carcinogen metabolizing enzymes and covalent DNA adduct formation in the exposed bronchial epithelium. Exposure of the bronchial epithelium to tobacco smoke appears to result in changes in tissue morphology, which are thought to be precursors of cancer. Most commonly lung cancers metastasize to pleura, brain, bone, pericardium, and liver. The decision to treat with surgery, radiation therapy, or chemotherapy is made on the basis of tumor histology, response to growth factors or hormones, and sensitivity to inhibitors or drugs. With current treatments, most patients die within one year of diagnosis. Earlier diagnosis and a systematic approach to identification, staging, and treatment of lung cancer could positively affect patient outcome.

Lung cancers progress through a series of morphologically distinct stages from hyperplasia to invasive carcinoma. Malignant lung cancers are divided into two groups comprising four

histopathological classes. The Non Small Cell Lung Carcinoma (NSCLC) group includes squamous cell carcinomas, adenocarcinomas, and large cell carcinomas and accounts for about 70% of all lung cancer cases. Adenocarcinomas typically arise in the peripheral airways and often form mucin secreting glands. Squamous cell carcinomas typically arise in proximal airways. The histogenesis of
5   squamous cell carcinomas may be related to chronic inflammation and injury to the bronchial epithelium, leading to squamous metaplasia. The Small Cell Lung Carcinoma (SCLC) group accounts for about 20% of lung cancer cases. SCLCs typically arise in proximal airways and exhibit a number of paraneoplastic syndromes including inappropriate production of adrenocorticotropin and anti-diuretic hormone.

10   Lung cancer cells accumulate numerous genetic lesions, many of which are associated with cytologically visible chromosomal aberrations. The high frequency of chromosomal deletions associated with lung cancer may reflect the role of multiple tumor suppressor loci in the etiology of this disease. Deletion of the short arm of chromosome 3 is found in over 90% of cases and represents one of the earliest genetic lesions leading to lung cancer. Deletions at chromosome arms 9p and 17p
15   are also common. Other frequently observed genetic lesions include overexpression of telomerase, activation of oncogenes such as K-ras and c-myc, and inactivation of tumor suppressor genes such as RB, p53 and CDKN2.

          Genes differentially regulated in lung cancer have been identified by a variety of methods. Using mRNA differential display technology, Manda, R. et al. (1999; Genomics 51:5-14) identified
20   five genes differentially expressed in lung cancer cell lines compared to normal bronchial epithelial cells. Among the known genes, pulmonary surfactant apoprotein A and alpha 2 macroglobulin were down regulated whereas nm23H1 was upregulated. Petersen, S. et al.. (2000; Int J. Cancer, 86:512-517) used suppression subtractive hybridization to identify 552 clones differentially expressed in lung tumor derived cell lines, 205 of which represented known genes. Among the known genes,
25   thrombospondin-1, fibronectin, intercellular adhesion molecule 1, and cytokeratins 6 and 18 were previously observed to be differentially expressed in lung cancers. Wang, T. et al. (2000; Oncogene 19:1519-1528) used a combination of microarray analysis and subtractive hybridization to identify 17 genes differentially overexpresssed in squamous cell carcinoma compared with normal lung epithelium. Among the known genes they identified were keratin isoform 6, KOC, SPRC, IGFb2,
30   connexin 26, plakofillin 1 and cytokeratin 13.
Breast Cancer

          There are more than 180,000 new cases of breast cancer diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (Gish, K. (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of

localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with

5   conventional screening methods to improve diagnosis and prognosis of this disease (Perou, C.M. et al. (2000) Nature 406:747-752).

Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, supra). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority

10   of breast cancer is due to non-inherited mutations that occur in breast epithelial cells.

The relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied. (See Khazaie, K. et al. (1993) Cancer and Metastasis Rev. 12:255-274, and references cited therein for a review of this area.) Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a

15   marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR

20   is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, S. S. et al. (1994) Am. J. Clin. Pathol. 102:S13-S24). Other known markers of breast cancer include a human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix Gla

25   protein which is overexpressed is human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down regulated in mammary carcinoma cells relative to normal mammary epithelial cells (Zhou, Z. et al. (1998) Int. J. Cancer 78:95-99; Chen, L. et al. (1990) Oncogene 5:1391-1395; Ulrix,

30   W. et al (1999) FEBS Lett 455:23-26; Sager, R. et al. (1996) Curr. Top. Microbiol. Immunol. 213:51-64; and Lee, S. W. et al. (1992) Proc. Natl. Acad. Sci. USA 89:2504-2508).

Cell lines derived from human mammary epithelial cells at various stages of breast cancer provide a useful model to study the process of malignant transformation and tumor progression as it has been shown that these cell lines retain many of the properties of their parental tumors for lengthy

culture periods (Wistuba, I.I. et al. (1998) Clin. Cancer Res. 4:2931-2938). Such a model is particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

Prostate Cancer

5   Prostate cancer is a common malignancy in men over the age of 50, and the incidence increases with age. In the US, there are approximately 132,000 newly diagnosed cases of prostate cancer and more than 33,000 deaths from the disorder each year.

      Once cancer cells arise in the prostate, they are stimulated by testosterone to a more rapid growth. Thus, removal of the testes can indirectly reduce both rapid growth and metastasis of the

10 cancer. Over 95 percent of prostatic cancers are adenocarcinomas which originate in the prostatic acini. The remaining 5 percent are divided between squamous cell and transitional cell carcinomas, both of which arise in the prostatic ducts or other parts of the prostate gland.

      As with most cancers, prostate cancer develops through a multistage progression ultimately resulting in an aggressive, metastatic phenotype. The initial step in tumor progression involves the

15 hyperproliferation of normal luminal and/or basal epithelial cells that become hyperplastic and evolve into early-stage tumors. The early-stage tumors are localized in the prostate but eventually may metastasize, particularly to the bone, brain or lung. About 80% of these tumors remain responsive to androgen treatment, an important hormone controlling the growth of prostate epithelial cells. However, in its most advanced state, cancer growth becomes androgen-independent and there is

20 currently no known treatment for this condition.

      A primary diagnostic marker for prostate cancer is prostate specific antigen (PSA). PSA is a tissue-specific serine protease almost exclusively produced by prostatic epithelial cells. The quantity of PSA correlates with the number and volume of the prostatic epithelial cells, and consequently, the levels of PSA are an excellent indicator of abnormal prostate growth. Men with prostate cancer

25 exhibit an early linear increase in PSA levels followed by an exponential increase prior to diagnosis. However, since PSA levels are also influenced by factors such as inflammation, androgen and other growth factors, some scientists maintain that changes in PSA levels are not useful in detecting individual cases of prostate cancer.

      Current areas of cancer research provide additional prospects for markers as well as potential

30 therapeutic targets for prostate cancer. Several growth factors have been shown to play a critical role in tumor development, growth, and progression. The growth factors Epidermal Growth Factor (EGF), Fibroblast Growth Factor (FGF), and Tumor Growth Factor alpha (TGFα) are important in the growth of normal as well as hyperproliferative prostate epithelial cells, particularly at early stages of tumor development and progression, and affect signaling pathways in these cells in various ways (Lin

J et al. (1999) Cancer Res. 59:2891-2897; Putz T et al. (1999) Cancer Res 59:227-233). The TGF-β family of growth factors are generally expressed at increased levels in human cancers and the high expression levels in many cases correlates with advanced stages of malignancy and poor survival (Gold LI (1999) Crit Rev Oncog 10:303-360). Finally, there are human cell lines representing both

5   the androgen-dependent stage of prostate cancer (LNCap) as well as the androgen-independent, hormone refractory stage of the disease (PC3 and DU-145) that have proved useful in studying gene expression patterns associated with the progression of prostate cancer, and the effects of cell treatments on these expressed genes (Chung TD (1999) Prostate 15:199-207).

Leukemias

10   Leukemias can be classified into four major categories, and all involve malignant transformation of pluripotent stem cells. Acute leukemias, both lymphoblastic (ALL) and myeloid (AML) types, are characterized by the presence of immature cells in the blood. Chronic leukemias, both lymphocytic (CLL) and myelocytic (CML), are associated with mature, differentiated cells, but proportions of each cell type are abnormal. For example, CLL patients usually have clonal expansion

15   of B cell lymphocytes. CML patients often have granulocytes of all stages of maturity present in blood, bone marrow, and other organs. Monoclonal antibodies specific for B- and T-cells are helpful diagnostic tools, in addition to histological analysis. Disease progresses as normal hematopoietic bone marrow is displaced by malignant cells. Cause has been determined to be genetic in some cases, and chemical or radiation-induced in others.

20   Aging and Senescence

Studies of the aging process or senescence have shown a number of characteristic cellular and molecular changes (Fauci et al. (1998) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, p.37). These characteristics include increases in chromosome structural abnormalities, DNA cross-linking, incidence of single-stranded breaks in DNA, losses in DNA methylation, and

25   degradation of telomere regions. In addition to these DNA changes, post-translational alterations of proteins increase including, deamidation, oxidation, cross-linking, and nonenzymatic glycation. Still further molecular changes occur in the mitochondria of aging cells through deterioration of structure. These changes eventually contribute to decreased function in every organ of the body.

Alzheimer's disease is a progressive neurodegenerative disorder that is characterized by the

30   formation of senile plaques and neurofibrillary tangles containing amyloid beta peptide. These plaques are found in limbic and association cortices of the brain, including hippocampus, temporal cortices, cingulate cortex, amygdala, nucleus basalis and locus caeruleus. Early in Alzheimer's pathology, physiological changes are visible in the cingulate cortex (Minoshima, S. et al. (1997) Annals of Neurology 42:85-94). In subjects with advanced Alzheimer's disease, accumulating

plaques damage the neuronal architecture in limbic areas and eventually cripple the memory process.

Genes Regulated in Dendritic Cell Differentiation

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression

5   of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for examining which genes are tissue specific, carrying out housekeeping functions, parts of a signaling cascade, or specifically related to a particular genetic predisposition, condition, disease, or disorder.

The potential application of gene expression profiling is particularly relevant to

10 characterizing lineage differences during cellular development that will improve diagnosis, prognosis, and treatment of disease. For example, both the levels and sequences expressed in dendritic cells from subjects with autoimmunity may be compared with the levels and sequences expressed in dendritic cells from normal subjects.

Dendritic cells (DC) are antigen presenting cells (APC) that play a key role in the primary

15 immune response because of their unique ability to present antigens to naive T cells. In addition, DC differentiate into separate subsets that sustain and regulate immune responses following initial contact with antigen. DC subsets include those that preferentially induce particular T helper 1 (Th1) or T helper 2 (Th2) responses and those that regulate B cell responses. Moreover, DC are increasingly being used to manipulate immune responses, either to downregulate an aberrant autoimmune response

20 or to enhance vaccination or a tumor-specific response.

DC are functionally specialized in correlation with their particular differentiation state. CD34+ myeloid cells found in the bone marrow mature in response to as yet unclear signals into CD14+ CD11c+ monocytes. An innate or antigen non-specific response takes place initially when monocytes circulate to nonlymphoid tissues and respond to lipopolysaccharide (LPS), a bacterially-

25 derived mitogen, and viruses. Such direct encounter with antigen causes secretion of pro-inflammatory cytokines that attract and regulate natural killer cells, macrophages, and eosinophils in the first line of defense against invading pathogens. Monocytes then mature into DC, which capture antigen highly efficiently through endocytosis and antigen receptor uptake. Antigen processing and presentation trigger activation and differentiation into mature DC that express MHC class II

30 molecules on the cell surface and efficiently activate T cells, initiating antigen-specific T cell and B cell responses. In turn, T cells activate DC through CD40 ligand - CD40 interactions, which stimulate expression of the costimulatory molecules CD80 and CD86, the latter most potent in amplifying T cell responses. DC interaction via CD40 with T cells also stimulates the production of inflammatory cytokines such as TNF alpha and IL-1. Engagement of RANK, a member of the TNF receptor family

by its ligand, TRANCE, which is expressed on activated T cells, enhances the survival of DC through inhibition of apoptosis, thereby enhancing T cell activation. The maturation and differentiation of monocytes into mature DC links the antigen non-specific innate immune response to the antigen-specific adaptive immune response.

5   The process by which monocytes differentiate into immature dendritic cells in vivo has not been fully elucidated. Incubation of monocytes with granulocyte-macrophage colony stimulating factor (GM-CSF) and interleukin (IL) -4 in vitro yields cells that exhibit functional and morphological characteristics equivalent to immature dendritic cells found in vivo. Moreover, incubation in vitro of immature dendritic cells with tumor necrosis factor alpha (TNF-α), CD40 ligand, LPS, or monocyte-
10   conditioned medium yields mature dendritic cells that are potent activators of naive T cells.

        The ability to manipulate DC in vitro and their capacity to mount an effective immune response with small numbers of DC and little antigen has led to potential immunotherapies for diseases such as cancer, AIDS, and infectious diseases; and enhancing vaccine efficacy. Spontaneous remissions of particular cancers such as renal cell carcinomas and melanomas indicate that the
15   immune system can respond to tumor antigens and eliminate tumors. However, tumors escape immune surveillance through a number of means including secretion of IL-10, macrophage colony stimulating factor, IL-6, and vascular endothelial growth factor, all of which inhibit DC activity and promote tolerance of tumor tissue. Delivery of tumor antigen-loaded DC to tumors can induce tumor-specific rejection in animal models. Similarly, pathogens can escape immune surveillance by altering
20   antigen processing and presentation pathways or interfering with maturation of antigen presenting cells. Rather than providing resistance, DC can complicate infection by hosting latent viruses such as Kaposi's virus and cytomegalovirus, complicating infection. HIV-1 and measles virus particles are efficiently produced in DC. Vaccines against tumors or infectious pathogens could be improved by systemic or local administration of DC loaded with tumor antigens or attenuated viral particles or
25   components, respectively.

        The expression of killer-inhibitor regulatory molecules, chemokines, chemokine receptors, and proteinases have been identified in DC through sequencing of ESTs. Continuing this search may reveal new lymphocyte-binding and antigen-processing molecules, transmembrane and secretory products, and transcription factors that may help to explain the specialized features of DC and allow
30   manipulation of the immune system.

Adipocyte maturation

        The primary function of adipose tissue is the ability to store and release fat during periods of feeding and fasting. White adipose tissue is the major energy reserve in periods of fasting, and its reserve is mobilized during energy deprivation. Adipose tissue is one of the primary target tissues for

insulin, and adipogenesis and insulin resistance are linked in type II diabetes, non-insulin dependent
diabetes mellitus (NIDDM). Cytologically the conversion of a preadipocytes into mature adipocytes
is characterized by deposition of fat droplets around the nuclei. The conversion process in vivo can
be induced by thiazolidinediones and other PPARγ agonists (Adams et al. (1997) J Clin Invest

5   100:3149-3153) which also lead to increased sensitivity to insulin and reduced plasma glucose and
blood pressure.

Pickup and Crook (1998; Diabetologia 41:1241-8) have suggested that NIDDM may result
from the inability of an individual with hypersensitive acute-phase immune response to carry out
normal cell signaling and repair. Steps in this process are highly correlated with long-term lifestyle

10   and environment and include: 1) high glucose stimulation of insulin and cytokine production, 2)
influence of various cytokines on tissue remodeling during adipocyte differentiation and their affect
on signaling pathways, and 3) occurrence of tissue damage when cytokines continue to be produced,
extracellular matrix components (ECM) are not recycled, and homeostasis is not timely restored.
Many cytokines and the receptors with which they interact are implicated in this process. These

15   cytokines include tumor necrosis factor, connective tissue growth factor, transforming growth factor-
beta, interleukin (IL)-13 and their receptors. Tumor necrosis factor contributes to insulin resistance
by inhibiting insulin-stimulated tyrosine phosphorylation of the insulin receptor. This, in turn,
prevents the insulin receptor from participating in normal signaling processes (Skolnik and
Marcusohn (1996) Cytokine Growth Factor Rev 7:161-173; Hotamisligil (1999) J Intern med

20   245:621-625). Connective tissue growth factor mediates the buildup of mesenglial matrix (Murphy et
al. (2000) J Biol Chem 274:5830-5834). Transforming growth factor-beta mediates the buildup of
mesenglial matrix of the kidney and affects vascular function through its interaction with the inositol
trisphosphate receptor, a key intracellular calcium channel (Sharma and McGowan (2000) Cytokine
Growth Factor Rev 11:115-123).

25   IL-13 and IL-4 are immuno-regulatory cytokines which share many overlapping biological
properties. They both promote growth of B-cells (McKenzie et al. (1993) Proc Natl Acad Sci
90:3735-3739), induce expression of germ line Cε transcripts, and direct naive B cells to switch to the
synthesis of IgE and IgG4 (Punnomen et al. (1993) Proc Natl Acad Sci 90:3730-3734). Similarly,
different isoforms of the IL-13 and IL-4 receptors interact to form four types of IL-13 receptor

30   complexes. In some instances, IL-13 utilizes a receptor complex composed of the IL-4 receptor-α
chain (Rα) and the IL-13Rα. Although the specific role of each chain in IL-13 signaling is unclear,
Ba/F3 cells transfected with IL-13Rα1 display a mitogenic response to IL-13, but cells transfected
with mouse IL-13Rα2 do not. In addition, a soluble IL-13Rα2/Fc fusion protein blocks the mitogenic
response to IL-13 (Donaldson et al. (1998) J Immunol 161:2317-2324). This suggests that IL-13Rα2

could serve as a dominant negative inhibitor or decoy receptor for IL-13. However, in colonic carcinoma cell lines, the receptor complex displayed growth inhibition which was associated with tyrosine phosphorylation of insulin receptor substrate-1. It is evident that more research is needed to establish 1) which isoforms of the receptor complex promote cell growth and which inhibit cell

5 growth and 2) whether this varies by cell or tissue type.

Most adipocyte research has been carried out using mouse cell lines. Recent evidence, however, indicates that culture conditions which stimulate mouse preadipocyte differentiation are different from those which induce human preadipocytes. In addition to the known genetic differences between these species, diploid human primary cells respond differently than aneuploid mouse cells.

10 Atherosclerosis and the associated coronary artery disease and cerebral stroke represent the most common cause of death in industrialized nations. Although certain key risk factors have been identified, a full molecular characterization that elucidates the causes and provide care for this complex disease has not been achieved. Molecular characterization of growth and regression of atherosclerotic vascular lesions requires identification of the genes that contribute to features of the

15 lesion including growth, stability, dissolution, rupture and, most lethally, induction of occlusive vessel thrombus. Vascular lesions principally involve the vascular endothelium and the surrounding smooth muscle tissue.

Development of atherosclerosis is understood to be induced by the presence of circulating lipoprotein. Lipoproteins, such as the cholesterol-rich low-density lipoprotein (LDL), accumulate in

20 the extracellular space of the vascular intima, and undergo modification. Oxidation of LDL (Ox-LDL) occurs most avidly in the sub-endothelial space where circulating antioxidant defenses are less effective. Mononuclear phagocytes enter the intima, differentiate into macrophages, and ingest modified lipids including Ox-LDL. During Ox-LDL uptake, macrophages produce cytokines (e.g. tumor necrosis factor α (TNF-α) and interleukin-1 (IL-1)) and growth factors (e.g. M-CSF, VEGF,

25 and PDGF-BB) that elicit further cellular events that modulate atherogenesis such as smooth muscle cell proliferation and production of extracellular matrix by vascular endothelium. Additionally, these macrophages may activate genes in endothelium and smooth muscle tissue involved in inflammation and tissue differentiation, including superoxide dismutatse (SOD), IL-8, and ICAM-1.

The vascular endothelium influences not only the three classically interacting components

30 of hemostasis: the vessel, the blood platelets and the clotting and fibrinolytic systems of plasma, but also the natural sequelae: inflammation and tissue repair. Two principal modes of endothelial behavior may be differentiated, best defined as an anti- and a prothrombotic state. Under physiological conditions endothelium mediates vascular dilatation (formation of nitric oxide (NO), PGI_2, adenosine, hyperpolarising factor), prevents platelet adhesion and activation (production of

adenosine, NO and PGI$_2$, removal of ADP), blocks thrombin formation (tissue factor pathway inhibitor, activation of protein C via thrombomodulin, activation of antithrombin III) and mitigates fibrin deposition (t- and scuplasminogen activator production). Adhesion and transmigration of inflammatory leukocytes are attenuated, e.g. by NO and IL-10, and oxygen radicals are efficiently
5   scavenged (urate, NO, glutathione, SOD).

When the endothelium is physically disrupted or functionally perturbed by postischemic reperfusion, acute and chronic inflammation, atherosclerosis, diabetes and chronic arterial hypertension, then completely opposing actions pertain. This prothrombotic, proinflammatory state is characterised by vaso-constriction, platelet and leukocyte activation and adhesion (externalisation,
10   expression and upregulation of, for example, von Willebrand factor, platelet activating factor, P-selectin, ICAM-1, IL-8, MCP-1, and TNF-α), promotion of thrombin formation, coagulation and fibrin deposition at the vascular wall (expression of tissue factor, PAI-1, and phosphatidyl serine) and, in platelet-leukocyte coaggregates, additional inflammatory interactions via attachment of platelet CD40-ligand to endothelial, monocyte and B-cell CD40. Since thrombin formation and inflammatory
15   stimulation set the stage for later tissue repair, complete abolition of such endothelial responses cannot be the goal of clinical interventions aimed at limiting procoagulatory, prothrombotic actions of a dysfunctional vascular endothelium. (See, e.g., Becker et al. (2000) Z Kardiol 89:160-167.)

Tumor necrosis factor α is a pleiotropic cytokine that a mediates immune regulation and inflammatory responses. TNF-α-related cytokines generate partially overlapping cellular responses,
20   including differentiation, proliferation, nuclear factor-κB (NF-κB) activation, and cell death, by triggering the aggregation of receptor monomers (Smith, C.A. et al. (1994) Cell 76:959-962). The cellular responses triggered by TNF-α are initiated through its interaction with distinct cell surface receptors (TNFRs). NF-κB is a transcription factor with a pivotal role in inducing genes involved in physiological processes as well as in the response to injury and infection. Activation of NF-κB
25   involves the phosphorylation and subsequent degradation of an inhibitory protein, IKB,and many of the proximal kinases and adaptor molecules involved in this process have been elucidated. Additionally, the NF-κB activation pathway from cell membrane to nucleus for IL-1 and TNF-α is now understood (Bowie and O'Neill (2000) Biochem Pharmacol 59:13-23).

Monocyte chemoattractant protein-1 (MCP-1) is known to play an important role in the
30   pathogenesis of atherosclerosis by inducing monocyte migration. TNF-α treatment of human umbilical vein endothelial cells (HUVECs) increased the cellular secretions of MCP-1 119-fold compared with untreated cells. Troglitazone, an insulin-sensitizing drug, significantly inhibited this TNF-α-induced increase in MCP-1 secretions and decreased mRNA levels (Ohta et al. (2000) Diabetes Res Clin Pract 48:171-176).

Treatment of confluent cultures of vascular smooth muscle cells (SMCs) with TNF-α

suppresses the incorporation of [³H]proline into both collagenase-digestible proteins (CDP) and

noncollagenous proteins (NCP). Such suppression by TNF-α is not observed in confluent bovine

aortic endothelial cells and human fibroblastic IMR-90 cells. TNF-α decreases the relative proportion

5   of collagen types IV and V suggesting that TNF-α modulates collagen synthesis by SMCs depending

on their cell density and therefore may modify formation of atherosclerotic lesions (Hiraga et al.

(2000) Life Sci 66:235-244).

Human coronary artery smooth muscle cells (CASMC) are primary cells isolated from the

tunica media (an intermediate muscular layer) of a human coronary artery. Vascular smooth muscle

10 cells are a model of increasing significance in vascular biology. It is now well known that besides

their obvious role in the regulation of vascular tone and, consequently, oxygen supply to various

tissues, their behavior under inflammatory conditions is an important factor in the development of

atherosclerosis and restenosis.

Human aortic endothelial cells (HAECs) are primary cells derived from the endothelium of a

15 human aorta. HAECs have been used as an experimental model for investigating *in vitro* the role of

the endothelium in human vascular biology. Activation of the vascular endothelium is considered to

be a central event in a wide range of both physiological and pathophysiological processes, such as

vascular tone regulation, coagulation and thrombosis, atherosclerosis, and inflammation.

Thus, vascular tissue genes differentially expressed during treatment of CASMC and HAEC

20 cell cultures with TNFα may reasonably be expected to be markers of the atherosclerotic process.

Biopharmacological tools for microarray analysis

Human umbilical vein endothelial cells (HUVECs) are a primary cell line derived from the

endothelium of the human umbilical vein. HUVECs are used to study the functional biology of

human endothelial cells *in vitro*. Activation of vascular endothelium is observed in physiological and

25 pathophysiological processes including vascular tone regulation, coagulation and thrombosis,

atherosclerosis, and inflammation.

Tumor necrosis factor-alpha (TNF-α) is a pleiotropic cytokine that plays a central role in

mediation of the inflammatory response through activation of multiple signal transduction pathways.

TNF-α is produced by activated lymphocytes, macrophages, and other white blood cells, and activates

30 endothelial cells.

PMA is an agonist of protein kinase C (PKC). PKC is a calcium-activated, phospholipid-

dependent serine- and threonine-specific kinase that, upon activation, phosphorylates a broad range of

secondary targets. TNF-α causes translocation of PKC from the cytosol to the membrane where it

phosphorylates a variety of targets.

Interleukin 1 beta (IL-1β)is a cytokine associated with acute inflammatory responses and is involved in processes such as fever induction, metabolic regulation, and bone remodeling. Both cells of the immune system (monocytes, dendritic cells, NK cells, platelets, and neutrophils) and somatic cells (osteoblasts, neurons, Schwann cells, oligodendrocytes, and adrenal cortical cells) can produce

5   IL-1β. IL-1β can induce its own production in monocytes, the production of adhesion molecules and chemokines in endothelial cells, and interferon gamma (IFN-γ) production by NK cells in conjunction with IL-12. IL-1 is produced as a single chain pro-molecule that must be cleaved by a specialized protease – IL-1 Converting Enzyme (ICE) – to acquire its function. Interleukin 10 (IL-10) is produced by CD4+ T cell clones and some CD8+ T cell clones. Human B cells, EBV-transformed

10   lymphoblastoid cell lines, and monocytes can also produce IL-10 upon activation. IL-10 is a pleiotrophic cytokine that can exert either immunostimulatory or immunosupressive effects on a variety of cell types. It is a potent immunosuppressant of macrophage functions. In vitro, IL-10 can inhibit the accessory function and antigen-presenting capacity of monocytes by, among other effects, downregulating class II MHC expression. Thus, IL-10 can inhibit monocyte/macrophage-dependent,

15   antigen specific proliferation of mouse Th1 clones as well as human Th0-, Th1-, and Th2- like T cells. IL-10 can also inhibit the monocyte/macrophage-dependent, antigen stimulated cytokine synthesis (especially IFN-g) by human PBMNC and NK. Additionally, IL-10 is a potent inhibitor of monocyte/macrophage activation and its resultant cytotoxic effects. It can suppress the production of numerous cytokines including TNF-a, IL-1, IL-6, and IL-10, as well as the synthesis of superoxide

20   anion, reactive oxygen intermediates, and reactive nitrogen intermediates by activated monocytes/ macrophages. As an immunostimulatory cytokine, IL-10 can act on B cells to enhance their viability, cell proliferation, Ig secretion, and class II MHC expression. Aside from B-lymphocytes, IL-10 is also a growth co-stimulator for thymocytes and mast cells, as well as an enhancer of cytotoxic T cell development.

25   Thiazolidinediones or peroxisome proliferator-activated receptor γ (PPAR-γ) agonists are a new class of antidiabetic agents that improve insulin sensitivity and reduce plasma glucose and blood pressure in subjects with type II diabetes. These agents can bind and activate an orphan nuclear receptor and some of them induce human adipocyte differentiation.

There is a need in the art for new compositions, including nucleic acids and proteins, for the

30   diagnosis, prevention, and treatment of cell proliferative disorders including cancer, developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, proteins associated with cell growth, differentiation, and death, referred to collectively as 'CGDD' and individually as 'CGDD-1,' 'CGDD-2,' 'CGDD-3,' 'CGDD-4,' 'CGDD-5,' 'CGDD-6,' 'CGDD-7,' 'CGDD-8,' 'CGDD-9,' 'CGDD-10,' 'CGDD-11,' 'CGDD-12,' 'CGDD-13,' 'CGDD-14,' 'CGDD-15,' 'CGDD-

5   16,' 'CGDD-17,' 'CGDD-18,' 'CGDD-19,' 'CGDD-20,' 'CGDD-21,' 'CGDD-22,' 'CGDD-23,' 'CGDD-24,' 'CGDD-25,' 'CGDD-26,' and 'CGDD-27 ' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified proteins associated with cell growth, differentiation, and death and/or their encoding polynucleotides for facilitating the drug

10   discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified proteins associated with cell growth, differentiation, and death and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a

15   polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide

20   having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-27.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the

25   group consisting of SEQ ID NO:1-27, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group

30   consisting of SEQ ID NO:1-27. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-27. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:28-54.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group

consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino

5   acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group

10   consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic

15   fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

20   Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active

25   fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group

30   consisting of SEQ ID NO:28-54, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40,

60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, b) a

5   polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides

10   comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments,

15   the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at

20   least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target

25   polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a polypeptide comprising a

30   naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and a pharmaceutically acceptable excipient.

In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional CGDD, comprising administering to a patient in need of such treatment the composition.

5   Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active

10   fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a

15   pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional CGDD, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an

20   amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino

25   acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional

30   CGDD, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an

amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) combining the

5   polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

        Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino

10   acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence

15   selected from the group consisting of SEQ ID NO:1-27. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide

20   in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

        Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, the method

25   comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

        Another embodiment provides a method for assessing toxicity of a test compound, said

30   method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at

least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target

5   polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, iii) a polynucleotide complementary to the polynucleotide of

10   i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above;.c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of

15   hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

20   Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and

25   domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

30   Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the

invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that
5   embodiments of the invention are not limited to the particular machines, instruments, materials, and
methods described, as these may vary. It is also to be understood that the terminology used herein is
for the purpose of describing particular embodiments only, and is not intended to limit the scope of
the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include
10   plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a
host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one
or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same
meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.
15   Although any machines, materials, and methods similar or equivalent to those described herein can be
used to practice or test the present invention, the preferred machines, materials and methods are now
described. All publications mentioned herein are cited for the purpose of describing and disclosing
the cell lines, protocols, reagents and vectors which are reported in the publications and which might
be used in connection with various embodiments of the invention. Nothing herein is to be construed
20   as an admission that the invention is not entitled to antedate such disclosure by virtue of prior
invention.

DEFINITIONS

"CGDD" refers to the amino acid sequences of substantially purified CGDD obtained from
any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and
25   human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of
CGDD. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other
compound or composition which modulates the activity of CGDD either by directly interacting with
CGDD or by acting on components of the biological pathway in which CGDD participates.

30   An "allelic variant" is an alternative form of the gene encoding CGDD. Allelic variants may
result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in
polypeptides whose structure or function may or may not be altered. A gene may have none, one, or
many allelic variants of its naturally occurring form. Common mutational changes which give rise to
allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides.

Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding CGDD include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CGDD or a

5   polypeptide with at least one functional characteristic of CGDD. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CGDD, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding CGDD. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of

10 amino acid residues which produce a silent change and result in a functionally equivalent CGDD. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CGDD is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids

15 may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a

20 polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

25 "Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of CGDD. Antagonists may include proteins such as antibodies, anticalins, nucleic acids,

30 carbohydrates, small molecules, or any other compound or composition which modulates the activity of CGDD either by directly interacting with CGDD or by acting on components of the biological pathway in which CGDD participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')$_2$, and Fv fragments, which are capable of binding an epitopic determinant.

Antibodies that bind CGDD polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired.

5   Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to

10   immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a

15   specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules.

20   The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH$_2$), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a

25   cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-

30   handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions

50

may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-

5   deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

10   The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic CGDD, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

15   "Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The

20   composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding CGDD or fragments of CGDD may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's

25   solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer

30   program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of

the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
5	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
10	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
15	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
20	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

25   Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the
30   absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative
35   polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or
40   absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a

diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the
5   evolution of new protein functions.

A "fragment" is a unique portion of CGDD or a polynucleotide encoding CGDD which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid
10   residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a
15   polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:28-54 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:28-54, for example, as distinct from any other sequence in the
20   genome from which the fragment was obtained. A fragment of SEQ ID NO:28-54 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:28-54 from related polynucleotides. The precise length of a fragment of SEQ ID NO:28-54 and the region of SEQ ID NO:28-54 to which the fragment corresponds are routinely determinable by one of ordinary skill in
25   the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-27 is encoded by a fragment of SEQ ID NO:28-54. A fragment of SEQ ID NO:1-27 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-27. For example, a fragment of SEQ ID NO:1-27 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-27. The precise
30   length of a fragment of SEQ ID NO:1-27 and the region of SEQ ID NO:1-27 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length"

polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

5   *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous

10   nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes

15   in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some

20   alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e

25   sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

30   Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

5   *Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence,
for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for
example, over the length of a fragment taken from a larger, defined polypeptide sequence, for
instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least

10   150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment
length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be
used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain
DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for

15   chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid
sequence in the non-antigen binding regions has been altered so that the antibody more closely
resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a

20   complementary strand through base pairing under defined hybridization conditions. Specific
hybridization is an indication that two nucleic acid sequences share a high degree of complementarity.
Specific hybridization complexes form under permissive annealing conditions and remain hybridized
after the "washing" step(s). The washing step(s) is particularly important in determining the
stringency of the hybridization process, with more stringent conditions allowing less non-specific

25   binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive
conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill
in the art and may be consistent among hybridization experiments, whereas wash conditions may be
varied among experiments to achieve the desired stringency, and therefore hybridization specificity.
Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about

30   1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature
under which the wash step is carried out. Such wash temperatures are typically selected to be about
5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic
strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of

the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

5   High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents

10   include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such

15   similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

        The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters,.

20   chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

        The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

        "Immune response" can refer to conditions associated with inflammation, trauma, immune

25   disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

        An "immunogenic fragment" is a polypeptide or oligopeptide fragment of CGDD which is capable of eliciting an immune response when introduced into a living organism, for example, a

30   mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of CGDD which is useful in any of the antibody production methods disclosed herein or known in the art.

        The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of CGDD. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other

5   biological, functional, or immunological properties of CGDD.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

10   "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

15   "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

20   "Post-translational modification" of an CGDD may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of CGDD.

"Probe" refers to nucleic acids encoding CGDD, their complements, or fragments thereof,

25   which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a

30   DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100,

or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

5   Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989; Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose

10   such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer

15   selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome

20   Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource

25   Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for

30   example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of

sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the

5 nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is

10 expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

15   "Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear

20 sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing CGDD, nucleic acids encoding CGDD, or fragments thereof may comprise a bodily fluid; an extract from a

25 cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular

30 structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are

removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides

5   by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

10   A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid

15   sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well

20   as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor

25   of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms

30   contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra.*

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at

5   least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of

10   polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between

15   individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having

20   at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater

25   sequence identity over a certain defined length of one of the polypeptides..

THE INVENTION

Various embodiments of the invention include new human proteins associated with cell growth, differentiation, and death (CGDD), the polynucleotides encoding CGDD, and the use of these

30   compositions for the diagnosis, treatment, or prevention of cell proliferative disorders including cancer, developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated

to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ

5   ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

10   Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the

15   nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

20   Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the

25   MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

            Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these

30   properties establish that the claimed polypeptides are proteins associated with cell growth, differentiation, and death.

            For example, SEQ ID NO:2 is 51% identical, from residue G11 to residue A548, to a human leucine-rich glioma-inactivated protein precursor (GenBank ID g4091819) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 4.3e-161,

which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a leucine-rich repeat domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.)

5   As another example, SEQ ID NO:3 is 90% identical, from residue M1 to residue L378, to Mus musculus Iroquois-class homeobox protein IRX2 (GenBank ID g9965418) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.5e-190, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains a homeobox domain as determined by searching for statistically

10   significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses and BLAST analyses of the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:3 is a homeodomain containing protein.

          As another example, SEQ ID NO:6 is 99% identical, from residue Q87 to residue T1723, to

15   human non-lens beta gamma-crystallin like protein (GenBank ID g2072425) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains a beta/gamma crystallin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of

20   conserved protein family domains. (See Table 3.) Data from MOTIFS, and other BLAST analyses provide further corroborative evidence that SEQ ID NO:6 is a beta gamma-crystallin like protein.

          As another example, SEQ ID NO:9 is 53% identical from residue M1 to residue K56, and 84% identical from residue G55 to residue G535, to Mus musculus SPAF (spermatogenesis associated factor, AAA family) (GenBank ID g4105619) as determined by the Basic Local Alignment

25   Search Tool (BLAST). (See Table 2.) The BLAST probability score is 6.4e-218, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 also contains an ATPase family associated with various cellular activities (AAA) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS

30   analyses provide further corroborative evidence that SEQ ID NO:9 is an AAA family protein.

          As another example, SEQ ID NO:14 is 85% identical, from residue M501 to residue I1440, to rat CPG2 (candidate plasticity-related genes) protein, which is involved among other things, in postnatal cortical development (Nedivi, E. et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93 (5), 2048-2053) (GenBank ID g1177322) as determined by the Basic Local Alignment Search Tool

(BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:14 also contains a calponin homology (CH) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See

5   Table 3.) Data from MOTIFS, PROFILESCAN, and fuuther BLAST analyses provide corroborative evidence that SEQ ID NO:14 is a CPG2 protein.

As another example, SEQ ID NO:17 is 95% identical, from residue M1 to residue D686, to murine metastasis associated protein 1 (GenBank ID g15077051) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which

10   indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:17 also contains BAH, ELM2, GATA zinc finger, and Myb-like DNA-binding domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST-PRODOM, BLAST-DOMO, and MOTIFS analyses provide further corroborative evidence that SEQ

15   ID NO:17 is a protein associated with cell growth, differentiation and death.

As another example, SEQ ID NO:21 is 97% identical, from residue M1 to residue F652, to human OS-9 precursor (GenBank ID g1322234) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from BLIMPS, BLAST,

20   and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:21 is a protein associated with cell growth, differentiation, and death.

As another example, SEQ ID NO:25 is 99% identical, from residue M1 to residue T533, to human OS-9 (GenBank ID g2780783) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.5e-290, which indicates the probability

25   of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:25 also has homology to proteins that are OS-9 proteins, as determined by BLAST analysis using the PROTEOME database. Data from BLAST analyses provide further corroborative evidence that SEQ ID NO:25 is an OS-9.

SEQ ID NO:1, SEQ ID NO:4-5, SEQ ID NO:7-8, SEQ ID NO:10-13, SEQ ID NO:15-16,

30   SEQ ID NO:18-20, SEQ ID NO:22-24, and SEQ ID NO:26-27 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-27 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of

these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or

5 genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:8-14 or that distinguish between SEQ ID NO:8-14 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA

10 libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from

15 the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as

20 FL_XXXXXX_N₁_N₂_YYYYY_N₃_N₄ represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N₁,₂,₃..., if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an

25 "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FLXXXXXX_gAAAAA_gBBBBB_1_N is a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank

30 protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The

following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses CGDD variants. A preferred CGDD variant is one which

has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the CGDD amino acid sequence, and which contains at least one functional or structural characteristic of CGDD.

Various embodiments also encompass polynucleotides which encode CGDD. In a particular

5 embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:28-54, which encodes CGDD. The polynucleotide sequences of SEQ ID NO:28-54, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

10 The invention also encompasses variants of a polynucleotide encoding CGDD. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding CGDD. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:28-54 which has at least about 70%, or

15 alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:28-54. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of CGDD.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant

20 of a polynucleotide encoding CGDD. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding CGDD, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence

25 identity to a polynucleotide encoding CGDD over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding CGDD. For example, a polynucleotide comprising a sequence of SEQ ID NO:51 and a polynucleotide comprising a sequence of SEQ ID NO:38 are splice variants of each other; and a

30 polynucleotide comprising a sequence of SEQ ID NO:52, a polynucleotide comprising a sequence of SEQ ID NO:53, a polynucleotide comprising a sequence of SEQ ID NO:54, and a polynucleotide comprising a sequence of SEQ ID NO:48 are splice variants of each other. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of CGDD.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CGDD, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide

5   sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CGDD, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode CGDD and its variants are generally capable of

10   hybridizing to polynucleotides encoding naturally occurring CGDD under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding CGDD or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular

15   codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CGDD and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode CGDD and

20   CGDD derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding CGDD or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of

25   hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:28-54 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

30   Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification

system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied

5   Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding CGDD may be extended utilizing a partial nucleotide sequence

10   and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a

15   circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations

20   may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon

25   junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been

30   size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze

the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary

sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-

specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the

emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate

5   software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire

process from loading of samples to computer analysis and electronic data display may be computer

controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments

which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode

10   CGDD may be cloned in recombinant DNA molecules that direct expression of CGDD, or fragments

or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the

genetic code, other polynucleotides which encode substantially the same or a functionally equivalent

polypeptides may be produced and used to express CGDD.

The polynucleotides of the invention can be engineered using methods generally known in the

15   art in order to alter CGDD-encoding sequences for a variety of purposes including, but not limited to,

modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by

random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be

used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed

mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation

20   patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such

as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No.

5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat.

Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or

25   improve the biological properties of CGDD, such as its biological or enzymatic activity or its ability

to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene

variants is produced using PCR-mediated recombination of gene fragments. The library is then

subjected to selection or screening procedures that identify those gene variants with the desired

properties. These preferred variants may then be pooled and further subjected to recursive rounds of

30   DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial"

breeding and rapid molecular evolution. For example, fragments of a single gene containing random

point mutations may be recombined, screened, and then reshuffled until the desired properties are

optimized. Alternatively, fragments of a given gene may be recombined with fragments of

homologous genes in the same gene family, either from the same or different species, thereby

maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding CGDD may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980)

5   Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, CGDD itself or a fragment thereof may be synthesized using chemical methods known . in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated

10   synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of CGDD, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid

15   chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, supra, pp. 28-53).

In order to express a biologically active CGDD, the polynucleotides encoding CGDD or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains

20   the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding CGDD. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding CGDD. Such signals

25   include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding CGDD and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should

30   be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression

vectors containing polynucleotides encoding CGDD and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel et al., *supra*, ch. 1, 3,

5   and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding CGDD. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression

10   vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945;

15   Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population

20   (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending

25   upon the use intended for polynucleotides encoding CGDD. For example, routine cloning, subcloning, and propagation of polynucleotides encoding CGDD can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding CGDD into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of

30   transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of CGDD are needed, e.g. for the production of antibodies, vectors which direct high level expression of CGDD may be used. For example, vectors containing

the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CGDD. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such

5   vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of CGDD. Transcription of polynucleotides

10   encoding CGDD may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs

15   can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding CGDD may be ligated

20   into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CGDD in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based

25   vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

30   For long term production of recombinant proteins in mammalian systems, stable expression of CGDD in cell lines is preferred. For example, polynucleotides encoding CGDD can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in

enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

5   Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk⁻* and *apr⁻* cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to

10   methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA

15   85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

20   Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CGDD is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding CGDD can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CGDD under the

25   control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding CGDD and that express CGDD may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and

30   protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CGDD using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and

fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CGDD is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect.

5   IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled

10   hybridization or PCR probes for detecting sequences related to polynucleotides encoding CGDD include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding CGDD, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA

15   polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

20   Host cells transformed with polynucleotides encoding CGDD may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CGDD may be designed to contain signal sequences which

25   direct secretion of CGDD through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves

30   a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding CGDD may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CGDD protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the

5   screening of peptide libraries for inhibitors of CGDD activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion

10   proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CGDD encoding sequence and the heterologous protein sequence,

15   so that CGDD may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled CGDD may be achieved *in vitro* using the

20   TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

CGDD, fragments of CGDD, or variants of CGDD may be used to screen for compounds that

25   specifically bind to CGDD. One or more test compounds may be screened for specific binding to CGDD. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to CGDD. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of CGDD can be used to screen for binding of test

30   compounds, such as antibodies, to CGDD, a variant of CGDD, or a combination of CGDD and/or one or more variants CGDD. In an embodiment, a variant of CGDD can be used to screen for compounds that bind to a variant of CGDD, but not to CGDD having the exact sequence of a sequence of SEQ ID NO:1-27. CGDD variants used to perform such screening can have a range of about 50% to about 99% sequence identity to CGDD, with various embodiments having 60%, 70%, 75%, 80%, 85%,

90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to CGDD can be closely related to the natural ligand of CGDD, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current

5   Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor CGDD (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to CGDD can be closely related to the natural receptor to which CGDD binds, at least a fragment of the receptor, or

10   a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for CGDD which is capable of propagating a signal, or a decoy receptor for CGDD which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques

15   include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different

20   specificities can be screened for specific binding to CGDD, fragments of CGDD, or variants of CGDD. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of CGDD. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of CGDD. In another embodiment, an antibody can be selected such that its binding specificity allows

25   for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of CGDD.

In an embodiment, anticalins can be screened for specific binding to CGDD, fragments of CGDD, or variants of CGDD. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184;

30   Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions

(e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit CGDD involves producing appropriate cells which express CGDD, either as a secreted protein or on

5   the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing CGDD or cell membrane fractions which contain CGDD are then contacted with a test compound and binding, stimulation, or inhibition of activity of either CGDD or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is

10   detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with CGDD, either in solution or affixed to a solid support, and detecting the binding of CGDD to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical

15   libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No.

20   6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991)

25   Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

CGDD, fragments of CGDD, or variants of CGDD may be used to screen for compounds that modulate the activity of CGDD. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for CGDD

30   activity, wherein CGDD is combined with at least one test compound, and the activity of CGDD in the presence of a test compound is compared with the activity of CGDD in the absence of the test compound. A change in the activity of CGDD in the presence of the test compound is indicative of a compound that modulates the activity of CGDD. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising CGDD under conditions suitable for CGDD activity, and the

assay is performed. In either of these assays, a test compound which modulates the activity of CGDD may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding CGDD or their mammalian homologs may

5   be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of

10   interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res.

15   25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

20   Polynucleotides encoding CGDD may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

25   Polynucleotides encoding CGDD can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding CGDD is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and

30   treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress CGDD, e.g., by secreting CGDD in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists

between regions of CGDD and proteins associated with cell growth, differentiation, and death. In addition, the expression of CGDD is closely associated with bone tumor, bronchial, brain, brain tumor, breast, endometrial, epithelial cell, lung, ovarian, ovary tumor, small intestine, spinal cord, and thyroid tissue, as well as breast fibrocystic disease, diseased cerebellum and aortic tissues,

5   unstimulated astrocyte cells, and a transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue. In addition, examples of tissues expressing CGDD can be found in Table 6 and can also be found in Example XI. Therefore, CGDD appears to play a role in cell proliferative disorders including cancer, developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta. In the treatment of disorders

10 associated with increased CGDD expression or activity, it is desirable to decrease the expression or activity of CGDD. In the treatment of disorders associated with decreased CGDD expression or activity, it is desirable to increase the expression or activity of CGDD.

Therefore, in one embodiment, CGDD or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or

15 activity of CGDD. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of

20 the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia,

25 genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a neurological

30 disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess,

suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis,

5   encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental

10   disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,

15   allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's

20   syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome,

25   complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune

30   disorders, ectopic pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease, galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puperty,

retrograde ejaculation and anejaculation, haemangioblastomas, cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumors; and a disorder of the placenta such as preeclampsia, choriocarcinoma, abruptio placentae, placenta previa, placental or maternal floor infarction, placenta accreta, increate, and percreta, extrachorial placentas,

5   chorangioma, chorangiosis, chronic villitis, placental villous endema, widespread fibrosis of the terminal villi, intervillous thrombi, hemorraghic endovasculitis, erythroblastosis fetalis, and nonimmune fetal hydrops.

In another embodiment, a vector capable of expressing CGDD or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased

10   expression or activity of CGDD including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified CGDD in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CGDD including, but not limited to, those provided above.

15   In still another embodiment, an agonist which modulates the activity of CGDD may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CGDD including, but not limited to, those listed above.

In a further embodiment, an antagonist of CGDD may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CGDD. Examples of such

20   disorders include, but are not limited to, those cell proliferative disorders including cancer, developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta described above. In one aspect, an antibody which specifically binds CGDD may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CGDD.

25   In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CGDD may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CGDD including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents.

30   Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CGDD may be produced using methods which are generally known in the art. In particular, purified CGDD may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CGDD. Antibodies to CGDD may also be generated using methods that are well known in the art. Such antibodies may include, but are

5   not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J.

10   Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with CGDD or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are

15   not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to

20   CGDD have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of CGDD amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

25   Monoclonal antibodies to CGDD may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al.

30   (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985)

Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CGDD-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R.
5   (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

10   Antibody fragments which contain specific binding sites for CGDD may also be generated. For example, such fragments include, but are not limited to, F(ab')$_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989)
15   Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CGDD and its
20   specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CGDD epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CGDD. Affinity is expressed as an
25   association constant, K_a, which is defined as the molar concentration of CGDD-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CGDD epitopes, represents the average affinity, or avidity, of the antibodies for CGDD. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a
30   particular CGDD epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the CGDD-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CGDD, preferably in active form, from the

antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

5   The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CGDD-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty,
10   supra; Coligan et al., supra).

In another embodiment of the invention, polynucleotides encoding CGDD, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding
15   CGDD. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CGDD (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered
20   intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271; Ausubel et al.,
25   supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding CGDD may be used for
30   somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475),

cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene
Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial
hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal,
R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii)

5   express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated
cell proliferation), or (iii) express a protein which affords protection against intracellular parasites
(e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D.
(1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399),
hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides*

10   *brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the
case where a genetic deficiency in CGDD expression or regulation causes disease, the expression of
CGDD from an appropriate population of transduced cells may alleviate the clinical manifestations
caused by the genetic deficiency.

   In a further embodiment of the invention, diseases or disorders caused by deficiencies in

15   CGDD are treated by constructing mammalian expression vectors encoding CGDD and introducing
these vectors by mechanical means into CGDD-deficient cells. Mechanical transfer technologies for
use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii)
ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene
transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev.

20   Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr.
Opin. Biotechnol. 9:445-450).

   Expression vectors that may be effective for the expression of CGDD include, but are not
limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors
(Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA),

25   and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). CGDD may
be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous
sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible
promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl.
Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and

30   H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid
(Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND;
Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter
(Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of
the endogenous gene encoding CGDD from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID
TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver
polynucleotides to target cells in culture and require minimal effort to optimize experimental
parameters. In the alternative, transformation is performed using the calcium phosphate method

5   (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al.
(1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of
these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with
respect to CGDD expression are treated by constructing a retrovirus vector consisting of (i) the

10   polynucleotide encoding CGDD under the control of an independent promoter or the retrovirus long
terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive
element (RRE) along with additional retrovirus cis-acting RNA sequences and coding sequences
required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are
commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc.

15   Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in
an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for
receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al.
(1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and
A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R.

20   et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining
retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant")
discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by
reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4$^+$ T-
cells), and the return of transduced cells to a patient are procedures well known to persons skilled in

25   the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-
7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716;
Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-
2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver

30   polynucleotides encoding CGDD to cells which have one or more genetic abnormalities with respect
to the expression of CGDD. The construction and packaging of adenovirus-based vectors are well
known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to
be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas
(Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are

described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

5   In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding CGDD to target cells which have one or more genetic abnormalities with respect to the expression of CGDD. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing CGDD to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has

10   been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the

15   control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of

20   the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

     In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding CGDD to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based

25   on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for CGDD into the

30   alphavirus genome in place of the capsid-coding region results in the production of a large number of CGDD-coding RNAs and the synthesis of high levels of CGDD in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy

application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of CGDD into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA

5   transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it

10   causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the

15   transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze

20   endonucleolytic cleavage of RNA molecules encoding CGDD.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for

25   secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically

30   synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA molecules encoding CGDD. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs

5   and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a

10 compound which is effective in altering expression of a polynucleotide encoding CGDD. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide

15 sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased CGDD expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding CGDD may be therapeutically useful, and in the treatment of disorders associated with decreased CGDD expression or activity, a compound which specifically promotes

20 expression of the polynucleotide encoding CGDD may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary

25 library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding CGDD is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted

30 biochemical system. Alterations in the expression of a polynucleotide encoding CGDD are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding CGDD. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without

exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression

5   system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide

10   sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient.

15   Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and

20   monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of

25   Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of CGDD, antibodies to CGDD, and mimetics, agonists, antagonists, or inhibitors of CGDD.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal,

30   enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and

proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

5   Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising CGDD or fragments thereof. For example, liposome preparations

10   containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, CGDD or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

15   For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

20   A therapeutically effective dose refers to that amount of active ingredient, for example CGDD or fragments thereof, antibodies of CGDD, and agonists, antagonists or inhibitors of CGDD, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose

25   lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity.

30   The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the

severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

5   Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, 10   conditions, locations, etc.

DIAGNOSTICS

          In another embodiment, antibodies which specifically bind CGDD may be used for the diagnosis of disorders characterized by expression of CGDD, or in assays to monitor patients being treated with CGDD or agonists, antagonists, or inhibitors of CGDD. Antibodies useful for diagnostic 15   purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CGDD include methods which utilize the antibody and a label to detect CGDD in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

20   A variety of protocols for measuring CGDD, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CGDD expression. Normal or standard values for CGDD expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to CGDD under conditions suitable for complex formation. The amount of standard complex formation may be 25   quantitated by various methods, such as photometric means. Quantities of CGDD expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

          In another embodiment of the invention, polynucleotides encoding CGDD may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, 30   complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of CGDD may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CGDD, and to monitor regulation of CGDD levels during therapeutic intervention.

          In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides,

including genomic sequences, encoding CGDD or closely related molecules may be used to identify nucleic acid sequences which encode CGDD. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe

5   identifies only naturally occurring sequences encoding CGDD, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the CGDD encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:28-54 or from genomic sequences including promoters, enhancers, and introns of the CGDD gene.

10   Means for producing specific hybridization probes for polynucleotides encoding CGDD include the cloning of polynucleotides encoding CGDD or CGDD derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a

15   variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding CGDD may be used for the diagnosis of disorders associated with expression of CGDD. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed

20   connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate,

25   salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such

30   as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal

disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural

muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating

diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess,

suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system

5   disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-

Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the

nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis,

encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central

nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic

10   nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other

neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, ·

inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental

disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD),

akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses,

15   postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration,

and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired

immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,

allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune

hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal

20   dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis,

dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins,

erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's

syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel

syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation,

25   osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid

arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus,

systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome,

complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal,

parasitic, protozoal, and helminthic infections, and trauma; a reproductive disorder such as a disorder

30   of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a

disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome,

ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune

disorders, ectopic pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease,

galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer

of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of
the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism,
pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puperty,
retrograde ejaculation and anejaculation, haemangioblastomas, cystsphaeochromocytomas,

5   paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumors; and a disorder of the
placenta such as preeclampsia, choriocarcinoma, abruptio placentae, placenta previa, placental or
maternal floor infarction, placenta accreta, increate, and percreta, extrachorial placentas,
chorangioma, chorangiosis, chronic villitis, placental villous endema, widespread fibrosis of the
terminal villi, intervillous thrombi, hemorraghic endovasculitis, erythroblastosis fetalis, and

10   nonimmune fetal hydrops. Polynucleotides encoding CGDD may be used in Southern or northern
analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and
multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect
altered CGDD expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, polynucleotides encoding CGDD may be used in assays that detect the
15   presence of associated disorders, particularly those mentioned above. Polynucleotides
complementary to sequences encoding CGDD may be labeled by standard methods and added to a
fluid or tissue sample from a patient under conditions suitable for the formation of hybridization
complexes. After a suitable incubation period, the sample is washed and the signal is quantified and
compared with a standard value. If the amount of signal in the patient sample is significantly altered

20   in comparison to a control sample then the presence of altered levels of polynucleotides encoding
CGDD in the sample indicates the presence of the associated disorder. Such assays may also be used
to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical
trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of
25   CGDD, a normal or standard profile for expression is established. This may be accomplished by
combining body fluids or cell extracts taken from normal subjects, either animal or human, with a
sequence, or a fragment thereof, encoding CGDD, under conditions suitable for hybridization or
amplification. Standard hybridization may be quantified by comparing the values obtained from
normal subjects with values from an experiment in which a known amount of a substantially purified
30   polynucleotide is used. Standard values obtained in this manner may be compared with values
obtained from samples from patients who are symptomatic for a disorder. Deviation from standard
values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated,
hybridization assays may be repeated on a regular basis to determine if the level of expression in the

patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or
5   overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

10   Additional diagnostic uses for oligonucleotides designed from the sequences encoding CGDD may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding CGDD, or a fragment of a polynucleotide complementary to the polynucleotide encoding CGDD, and will be employed under optimized conditions for identification of a specific gene or
15   condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding CGDD may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods
20   of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding CGDD are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary
25   structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual
30   overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding

5   lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in

10 diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

15   Methods which may also be used to quantify the expression of CGDD include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is

20 presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers

25 of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a

30 pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, CGDD, fragments of CGDD, or antibodies specific for CGDD may be used as elements on a microarray. The microarray may be used to monitor or measure protein-

protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by

5   quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the

10   hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*,

15   as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed

20   molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number

25   of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids

30   in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present

5   invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression

10   in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using

15   two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot

20   is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed

25   by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for CGDD to quantify the levels of CGDD expression. In one embodiment, the antibodies are used as elements on a

30   microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at

each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson,

5   N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

10   In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic

15   response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are

20   incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

25   Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well

30   known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding CGDD may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be

preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs),

5   yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome

10   region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM)

15   World Wide Web site. Correlation between the location of the gene encoding CGDD on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts. :

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps.

20   Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23,

25   any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CGDD, its catalytic or immunogenic fragments, or

30   oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CGDD and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds

having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application
WO84/03564). In this method, large numbers of different small test compounds are synthesized on a
solid substrate. The test compounds are reacted with CGDD, or fragments thereof, and washed.
Bound CGDD is then detected by methods well known in the art. Purified CGDD can also be coated
5   directly onto plates for use in the aforementioned drug screening techniques. Alternatively,
non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing
antibodies capable of binding CGDD specifically compete with a test compound for binding CGDD.
In this manner, antibodies can be used to detect the presence of any peptide which shares one or more
10   antigenic determinants with CGDD.

In additional embodiments, the nucleotide sequences which encode CGDD may be used in
any molecular biology techniques that have yet to be developed, provided the new techniques rely on
properties of nucleotide sequences that are currently known, including, but not limited to, such
properties as the triplet genetic code and specific base pair interactions.

15   Without further elaboration, it is believed that one skilled in the art can, using the preceding
description, utilize the present invention to its fullest extent. The following embodiments are,
therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure
in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below,
20   including U.S. Ser. No. 60/311,017, U.S. Ser. No. 60/313,070, U.S. Ser. No. 60/313,071, U.S. Ser.
No. 60/314,678, U.S. Ser. No. 60/316,692, U.S. Ser. No. 60/317,913, U.S. Ser. No. 60/322,182, U.S.
Ser. No. 60/342,761, U.S. Ser. No. 60/340,747, and U.S. Ser. No. 60/369,129, are hereby expressly
incorporated by reference.

25   **EXAMPLES**

I.   Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database
(Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium
isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of
30   denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine
isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with
chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and
ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA

purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA

5   purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription

10   was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites

15   of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-

20   BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II.   Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an

25   AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a

30   high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III.   Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows.
Sequencing reactions were processed using standard methods or high-throughput instrumentation
such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal
5   cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the
MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared
using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the
ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).
Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides
10   were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the
ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI
protocols and base calling software; or other sequence analysis systems known in the art. Reading
frames within the cDNA sequences were identified using standard methods (Ausubel et al., *supra*, ch.
7). Some of the cDNA sequences were selected for extension using the techniques disclosed in
15   Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing
vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and
programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The
Incyte cDNA sequences or translations thereof were then queried against a selection of public
20   databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and
BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*,
Rattus norvegicus, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*,
Schizosaccharomyces pombe, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden
Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft,
25   D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as
SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002)
Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus
primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol.
6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and
30   HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide
sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences,
or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA
assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and
Consed, and cDNA assemblages were screened for open reading frames using programs based on

GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases
5   (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using
10 default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and
15 threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value,
20 the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:28-54. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

25 **IV.   Identification and Editing of Coding Sequences from Genomic DNA**

Putative proteins associated with cell growth, differentiation, and death were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J.
30 Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode proteins associated

with cell growth, differentiation, and death, the encoded polypeptides were analyzed by querying against PFAM models for proteins associated with cell growth, differentiation, and death. Potential proteins associated with cell growth, differentiation, and death were also identified by homology to Incyte cDNA sequences that had been annotated as proteins associated with cell growth,

5   differentiation, and death. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing

10   evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted

15   coding sequences.

V.   Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example

20   III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on

25   more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear

30   along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan

were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of CGDD Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:28-54 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:28-54 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified

disease genes map within or in proximity to the intervals indicated above.

VII.   Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs

5   from a particular cell type or tissue have been bound (Sambrook, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the

10   computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \text{ x minimum } \{\text{length(Seq. 1), length(Seq. 2)}\}}$$

15

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is

20   calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the

25   entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding CGDD are analyzed with respect to the tissue

30   sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male;

35   germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas;

respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma,

5   cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding CGDD. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII.   Extension of CGDD Encoding Polynucleotides

10   Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in

15   length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

         Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

20   High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair

25   PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

30   The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the

concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

5   The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham

10 Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase

15 (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20%

20 dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for

25 such extension, and an appropriate genomic library.

IX.   Identification of Single Nucleotide Polymorphisms in CGDD Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:28-54 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the

30 identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original

chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated
algorithms to identify errors introduced during laboratory processing, such as those caused by reverse
transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated
algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to
5   contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in
immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high
throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in
four different human populations. The Caucasian population comprised 92 individuals (46 male, 46
10   female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The
African population comprised 194 individuals (97 male, 97 female), all African Americans. The
Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The
Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown
of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele
15   frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed
no allelic variance in this population were not further tested in the other three populations.

X.   Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:28-54 are employed to screen cDNAs,
genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base
20   pairs, is specifically described, essentially the same procedure is used with larger nucleotide
fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06
software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of
[γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN,
Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25
25   superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7
counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of
human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I,
Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon
30   membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16
hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature
under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate.
Hybridization patterns are visualized using autoradiography or an alternative imaging means and
compared.

XI.   Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the

5   aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may

10   be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may

15   comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a

20   fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

25   ## Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM

30   dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20

minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is

5   then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification

10   uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water

15   washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S.

20   Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water.

25   Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and

30   Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for

about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an

5   Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a

10 resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The

15 emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on

20 the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the

25 two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high

30 signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then

integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements that exhibited at least about a two-fold change in expression, a signal-to-background ratio of at least 2.5, and an element spot size of at least 40% were identified as differentially

5   expressed using the GEMTOOLS program (Incyte Genomics).

<u>Expression</u>

        For example, SEQ ID NO:28 was differentially expressed in cancer cells versus normal cells based on microarray experimentation. SEQ ID NO:28 showed differential expression as determined by microarray analysis. Histological and molecular evaluation of breast tumors reveals that the

10   development of breast cancer evolves through a multi-step process whereby pre-malignant mammary epithelial cells undergo a relatively defined sequence of events leading to tumor formation. A cross-comparison experimental design was used to evaluate the expression of cDNAs from two human breast tumor cell lines (MCF-7 and BT20) at various stages of tumor progression, as compared to a non-malignant mammary epithelial cell line, HMEC (Clonetics, San Diego, CA). All cell cultures

15   were propagated in media according to the supplier's recommendations and grown to 70-80% confluence prior to RNA isolation.

        The expression of SEQ ID NO:28 was decreased at least two-fold in MCF-7 cells, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year old female. MCF-7 has retained characteristics of the mammary epithelium such as the ability to process

20   estradiol via cytoplasmic estrogen receptors and the capacity to form domes in culture. The expression of SEQ ID NO:28 was also found to be decreased by at least two-fold in BT20 cells, a breast carcinoma cell line derived *in vitro* from cells emigrating out of thin slices of the tumor mass isolated from a 74-year old female. These experiments indicate that SEQ ID NO:28 was significantly underexpressed in the breast tumor cell lines tested, further establishing the utility of SEQ ID NO:1

25   and SEQ ID NO:28 as diagnostic markers or as therapeutic targets for breast cancer.

        For example, SEQ ID NO:30 showed differential expression in lung tissue from patients with lung cancer compared to matched microscopically normal tissues from the same donors as determined by microarray analysis. The expression of CGDD-3 was decreased at least two-fold in cancerous lung tissue. In an alternative example, SEQ ID NO:30 showed differential expression in breast tissue from

30   a patient with breast cancer compared to matched microscopically normal tissue from the same donor as determined by microarray analysis. The expression of CGDD-3 was increased at least two-fold in cancerous breast tissue. SEQ ID NO:30 also showed differential expression in brain cingulate from a patient with Alzheimer's disease compared to matched microscopically normal tissue from the same donor as determined by microarray analysis. The expression of CGDD-3 was increased at least two-

fold in cingulate tissue with Alzheimer's disease. Therefore, SEQ ID NO:30 is useful in disease

staging and in diagnostic assays for cell proliferative disorders, including lung cancer and breast

cancer, and for neurological disorders, including Alzheimer's disease.

⟩   SEQ ID NO:33 showed differential expression as determined by microarray analysis. BT-20

5   cells are a breast carcinoma cell line derived *in vitro* from cells which emigrated out of thin slices of a

tumor mass isolated from a 74-year-old female. The treatment of BT-20 cells with epidermal growth

factor (EGF), a potent mitogen, revealed that the expression of SEQ ID NO:33 was down-regulated at

least two-fold. Additionally, a comparison of a primary breast epithelial cell line (HMEC cells)

isolated from a normal donor with BT-20, MCF7 (a nonmalignant breast adenocarcinoma cell line

10   isolated from the pleural effusion of a 69-year old female), MDA-mb-231 (a breast tumor cell line

isolated from the pleural effusion of a 51-year-old female which forms poorly differentiated

adenocarcinoma in nude mice), Sk-BR-3 (a breast adenocarcinoma cell line isolated from a malignant

pleural effusion of a 43-year-old female), and T-47D (a breast carcinoma cell line obtained from a 54-

year-old female with an infiltrating ductal carcinoma of the breast) cells also indicated that SEQ ID

15   NO:33 was down-regulated at least two-fold in each of the cancerous breast cell lines when compared

to the HMEC cell line. Furthermore, the expression of SEQ ID NO:33 was also found to be down-

regulated in two out of three prostate tumor cell lines when compared to normal donor primary

prostate epithelial cells. These experiments indicate that SEQ ID NO:33, encoding SEQ ID NO:6 was

significantly underexpressed in the breast tumor and prostate tumor cell lines tested, further

20   establishing the utility of SEQ ID NO:6 and SEQ ID NO:33 for disease staging, as diagnostic

markers, or as therapeutic targets for both breast and prostate cancer.

In another example, SEQ ID NO:42 showed differential expression as determined by

microarray analysis. Histological and molecular evaluation of breast tumors reveals that the

development of breast cancer evolves through a multi-step process whereby pre-malignant mammary

25   epithelial cells undergo a relatively defined sequence of events leading to tumor formation. An early

event in tumor development is ductal hyperplasia. Cells undergoing rapid neoplastic growth

gradually progress to invasive carcinoma and become metastatic to the lung, bone, and potentially

other organs. Several variables that may influence the process of tumor progression and malignant

transformation include genetic factors, environmental factors, growth factors, and hormones. Based

30   on the complexity of this process, it is critical to study a population of human mammary epithelial

cells undergoing the process of malignant transformation, and to associate specific stages of

progression with phenotypic and molecular characteristics.

In this experiment, primary human breast epithelial cells (HMECs) were compared to breast

carcinoma cell lines at various stages of tumor progression. All cells were grown in the supplier's

recommended medium to 70-80% confluence prior to RNA harvest. MCF7 is a nonmalignant breast adenocarcinoma cell line isolated from the pleura effusion of a 69-year old female. MCF7 has retained characertistics of the mammary epithelium such as the ability to process estradiol via cytoplasmic estrogen receptors and the capacity to form domes in culture. T-47D is a breast

5   carcinoma cell line isolated from a pleural effusion obtained from a 54-year old female with infiltrating ductal carcinoma of the breast. Sk-BR-3 is a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year old female. It forms poorly differentiated adenocarcinoma when injected into nude mice. The gene expression of SEQ ID NO:42 was decreased by at least two-fold in the breast cancer lines as compared to the HMEC cells.

10   In another experiment, the gene expression profile of SEQ ID NO:42 was evaluated in various breast cancer cell lines as compared to a nonmalignant breast epithelial cell line, MCF-10A (a breast mammary gland cell line, e.g. with luminal ductal characteristics, that was isolated from a 36-year old female with fibrocystic breast disease). The cells were grown in defined serum-free TCH medium to 70-80% confluence prior to RNA harvest. The expression of SEQ ID NO:42 was decreased by at

15   least two-fold in the following cell lines: MCF7, and BT-20, a breast carcinoma cell line derived *in vitro* from the cells emigrating out of thin slices from the tumor mass isolated from a 74-year old female.

        In another example, the gene expression profile of SEQ ID NO:42 was evaluated in various breast cancer lines as compared to MCF-10A cells, all of which were grown in defined serum-free

20   H14 medium to 70-80% confluence prior to RNA harvest. Under these conditions, the expression of SEQ ID NO:42 was decreased by at least two-fold in T-47D cells as compared to MCF-10A cells.

        In another example, SEQ ID NO:45 was differentially expressed in response to treatment with steroids. Steroids are a class of lipid-soluble molecules, including cholesterol, bile acids, vitamin D, and hormones, that share a common four-ring structure based on cyclopentanoperhydro-

25   phenanthrene and that carry out a wide variety of functions. Cholesterol, for example, is a component of cell membranes that controls membrane fluidity. It is also a precursor for bile acids which solubilize lipids and facilitate absorption in the small intestine during digestion. Vitamin D regulates the absorption of calcium in the small intestine and controls the concentration of calcium in plasma. Steroid hormones, produced by the adrenal cortex, ovaries, and testes, include

30   glucocorticoids, mineralocorticoids, androgens, and estrogens. They control various biological processes by binding to intracellular receptors that regulate transcription of specific genes in the nucleus. Glucocorticoids, for example, increase blood glucose concentrations by regulation of gluconeogenesis in the liver, increase blood concentrations of fatty acids by promoting lipolysis in adipose tissues, modulate sensitivity to catcholamines in the central nervous system, and reduce

inflammation. The principal mineralocorticoid, aldosterone, is produced by the adrenal cortex and acts on cells of the distal tubules of the kidney to enhance sodium ion reabsorption. Androgens, produced by the interstitial cells of Leydig in the testis, include the male sex hormone testosterone, which triggers changes at puberty, the production of sperm and maintenance of secondary sexual

5   characteristics. Female sex hormones, estrogen and progesterone, are produced by the ovaries and also by the placenta and adrenal cortex of the fetus during pregnancy. Estrogen regulates female reproductive processes and secondary sexual characteristics. Progesterone regulates changes in the endometrium during the menstrual cycle and pregnancy.

        Steroid hormones are widely used for fertility control and in anti-inflammatory treatments for

10   physical injuries and diseases such as arthritis, asthma, and auto-immune disorders. Progesterone, a naturally occurring progestin, is primarily used to treat amenorrhea, abnormal uterine bleeding, or as a contraceptive. Endogenous progesterone is responsible for inducing secretory activity in the endometrium of the estrogen-primed uterus in preparation for the implantation of a fertilized egg and for the maintenance of pregnancy. It is secreted from the corpus luteum in response to luteinizing

15   hormone (LH). The primary contraceptive effect of exogenous progestins involves the suppression of the midcycle surge of LH. At the cellular level, progestins diffuse freely into target cells and bind to the progesterone receptor. Target cells include the female reproductive tract, the mammary gland, the hypothalamus, and the pituitary. Once bound to the receptor, progestins slow the frequency of release of gonadotropin releasing hormone from the hypothalamus and blunt the pre-ovulatory LH

20   surge, thereby preventing follicular maturation and ovulation. Progesterone has minimal estrogenic and androgenic activity. Progesterone is metabolized hepatically to pregnanediol and conjugated with glucuronic acid.

        Medroxyprogesterone (MAH), also known as 6α-methyl-17-hydroxyprogesterone, is a synthetic progestin with a pharmacological activity about 15 times greater than progesterone. MAH

25   is used for the treatment of renal and endometrial carcinomas, amenorrhea, abnormal uterine bleeding, and endometriosis associated with hormonal imbalance. MAH has a stimulatory effect on respiratory centers and has been used in cases of low blood oxygenation caused by sleep apnea, chronic obstructive pulmonary disease, or hypercapnia.

        Corticosteroids are used to relieve inflammation and to suppress the immune response. They

30   inhibit eosinophil, basophil, and airway epithelial cell function by regulation of cytokines that mediate the inflammatory response. They inhibit leukocyte infiltration at the site of inflammation, interfere in the function of mediators of the inflammatory response, and suppress the humoral immune response. Corticosteroids are used to treat allergies, asthma, arthritis, and skin conditions. Beclomethasone is a synthetic glucocorticoid that is used to treat steroid-dependent asthma, to relieve

symptoms associated with allergic or nonallergic (vasomotor) rhinitis, or to prevent recurrent nasal polyps following surgical removal. The anti-inflammatory and vasoconstrictive effects of intranasal beclomethasone are 5000 times greater than those produced by hydrocortisone. Budesonide is a corticosteroid used to control symptoms associated with allergic rhinitis or asthma. Budesonide has

5   high topical anti-inflammatory activity but low systemic activity. Dexamethasone is a synthetic glucocorticoid used in anti-inflammatory or immunosuppressive compositions. It is also used in inhalants to prevent symptoms of asthma. Due to its greater ability to reach the central nervous system, dexamethasone is usually the treatment of choice to control cerebral edema. Dexamethasone is approximately 20-30 times more potent than hydrocortisone and 5-7 times more potent than

10   prednisone. Prednisone is metabolized in the liver to its active form, prednisolone, a glucocorticoid with anti-inflammatory properties. Prednisone is approximately 4 times more potent than hydrocortisone and the duration of action of prednisone is intermediate between hydrocortisone and dexamethasone. Prednisone is used to treat allograft rejection, asthma, systemic lupus erythematosus, arthritis, ulcerative colitis, and other inflammatory conditions. Betamethasone is a synthetic

15   glucocorticoid with antiinflammatory and immunosuppressive activity and is used to treat psoriasis and fungal infections, such as athlete's foot and ringworm.

The anti-inflammatory actions of corticosteroids are thought to involve phospholipase A_2 inhibitory proteins, collectively called lipocortins. Lipocortins, in turn, control the biosynthesis of potent mediators of inflammation such as prostaglandins and leukotrienes by inhibiting the release of

20   the precursor molecule arachidonic acid. Proposed mechanisms of action include decreased IgE synthesis, increased number of β-adrenergic receptors on leukocytes, and decreased arachidonic acid metabolism. During an immediate allergic reaction, such as in chronic bronchial asthma, allergens bridge the IgE antibodies on the surface of mast cells, which triggers these cells to release chemotactic substances. Mast cell influx and activation, therefore, is partially responsible for the

25   inflammation and hyperirritability of the oral mucosa in asthmatic patients. This inflammation can be retarded by administration of corticosteroids.

The effects upon liver metabolism and hormone clearance mechanisms are important to understand the pharmacodynamics of a drug. The human C3A cell line is a clonal derivative of HepG2/C3 (hepatoma cell line, isolated from a 15-year-old male with liver tumor), which was

30   selected for strong contact inhibition of growth. The use of a clonal population enhances the reproducibility of the cells. C3A cells have many characteristics of primary human hepatocytes in culture: i) expression of insulin receptor and insulin-like growth factor II receptor; ii) secretion of a high ratio of serum albumin compared with α-fetoprotein iii) conversion of ammonia to urea and glutamine; iv) metabolize aromatic amino acids; and v) proliferate in glucose-free and insulin-free

medium. The C3A cell line is now well established as an in vitro model of the mature human liver (Mickelson et al. (1995) Hepatology 22:866-875; Nagendra et al. (1997) Am J Physiol 272:G408-G416).

The expression of SEQ ID NO:45 was increased by at least two-fold in drug-treated, as

5 compared to untreated, C3A cells, under the following conditions: (1) at a dose of 10 μM medroxyprogesterone (MAH), for a duration of 1 and 3 hours of treatment; at a dose of 1 μM beclomethasone for a duration of 3 hours of treatment, and 10 μM for a duration of 1 and 3 hours of treatment; at doses of 1 and 10 μM budesonide for a duration of 1 and 3 hours of treatment, and at a dose of 100 μM for a duration of 1 and 6 hours of treatment; and lastly, at a dose of 10 μM

10 betamethasone for a duration of 1 hour of treatment, and at a dose of 100 μM for a duration of 1, 3, and 6 hours of treatment.

These experiments demonstrate a significant differential expression profile for SEQ ID NO:42 and SEQ ID NO:45, and further establish their utility as diagnostic markers, tools for disease staging, or as therapeutic targets for diseases in which proteins associated with cell growth,

15 differentiation, or death are involved.

In another example, SEQ ID NO:49 was differentially expressed using microarray techniques. Lung cancers are divided into four histopathologically distinct groups. Three groups (squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) are classified as non-small cell lung cancers (NSCLCs). The fourth group of cancers is referred to as small cell lung cancer (SCLC). Collectively,

20 NSCLCs account for approximately 70% of cases while SCLCs account for approximately 18% of cases. The molecular and cellular biology underlying the development and progression of lung caner are incompletely understood. Deletions on chromosome 3 are common in this disease and are thought to indicate the presence of a tumor suppressor gene in this region. Activating mutations in K-ras are commonly found in lung cancer and are the basis of one of the mouse models for the disease. Various

25 experiments were performed in which normal lung was compared to lung tumor from the same donor. The expression of SEQ ID NO:49 was decreased by at least two-fold in the lung tumor samples.

The majority of ovarian cancers are derived from epithelial cells, and 70% of patients with epithelial ovarian cancers present with late-stage disease. As a result, the long-term survival rates for this disease are very low. Identification of early-stage markers for ovarian cancer would significantly

30 increase the survival rate. The molecular events that lead to ovarian cancer are poorly understood. Some of the known aberrations include mutation of p53 and microsatellite instability. Since gene expression patterns likely vary when normal ovary is compared to ovarian tumors, we have examined gene expression in these tissues to identify possible markers for ovarian cancer. The expression of SEQ ID NO:49 was increased by at least two-fold in ovarian tumor tissue as compared to normal

ovary tissue from the same donor.

These experiments indicate that a significant differential expression of SEQ ID NO:22 under various conditions further establishes the utility of SEQ ID NO:49 as a diagnostic marker, a tool for disease staging, or as a therapeutic target for cancer.

5 **XII.   Complementary Polynucleotides**

Sequences complementary to the CGDD-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CGDD. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are

10 designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CGDD. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CGDD-encoding transcript.

15 **XIII.   Expression of CGDD**

Expression and purification of CGDD is achieved using bacterial or virus-based expression systems. For expression of CGDD in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid

20 promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CGDD upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CGDD in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus

25 (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CGDD by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases.

30 Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, CGDD is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step,

affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from CGDD at

5 specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified CGDD obtained by these methods can be used directly in the assays shown

10 in Examples XVII, XVIII, and XIX, where applicable.

XIV.   Functional Assays

CGDD function is assessed by expressing the sequences encoding CGDD at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice

15 include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to

20 distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the

25 uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity

30 with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of CGDD on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CGDD and either CD64 or CD64-GFP. CD64 and

CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human

immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using

magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success

NY). mRNA can be purified from the cells using methods well known by those of skill in the art.

5   Expression of mRNA encoding CGDD and other genes of interest can be analyzed by northern

analysis or microarray techniques.

XV.   Production of CGDD Specific Antibodies

CGDD substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g.,

Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to

10   immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the CGDD amino acid sequence is analyzed using LASERGENE software

(DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is

synthesized and used to raise antibodies by means known to those of skill in the art. Methods for ·

selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well

15   described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A

peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-

Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to

increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH

20   complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CGDD

activity by, for example, binding the peptide or CGDD to a substrate, blocking with 1% BSA, reacting

with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI.   Purification of Naturally Occurring CGDD Using Specific Antibodies

Naturally occurring or recombinant CGDD is substantially purified by immunoaffinity   ·

25   chromatography using antibodies specific for CGDD. An immunoaffinity column is constructed by

covalently coupling anti-CGDD antibody to an activated chromatographic resin, such as

CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and

washed according to the manufacturer's instructions.

Media containing CGDD are passed over the immunoaffinity column, and the column is

30   washed under conditions that allow the preferential absorbance of CGDD (e.g., high ionic strength

buffers in the presence of detergent). The column is eluted under conditions that disrupt

antibody/CGDD binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such

as urea or thiocyanate ion), and CGDD is collected.

XVII. Identification of Molecules Which Interact with CGDD

CGDD, or biologically active fragments thereof, are labeled with [125]I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CGDD, washed, and any wells with labeled CGDD complex are assayed. Data obtained using different concentrations of

5   CGDD are used to calculate values for the number, affinity, and association of CGDD with the candidate molecules.

Alternatively, molecules interacting with CGDD are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

10   CGDD may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of CGDD Activity

15   CGDD activity is demonstrated by measuring the induction of terminal differentiation or cell cycle progression when CGDD is expressed at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies, Gaithersburg, MD) and PCR 3.1 (Invitrogen, Carlsbad, CA), both of which contain the

20   cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector.

25   Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP) (Clontech, Palo Alto, CA), CD64, or a CD64-GFP fusion protein. Flow cytometry detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell cycle progression or terminal differentiation. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward

30   light scatter and 90 degree side light scatter; up or down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford,

New York, NY.

Alternatively, an *in vitro* assay for CGDD activity measures the transformation of normal human fibroblast cells overexpressing antisense CGDD RNA (Garkavtsev, I. and K. Riabowol (1997) Mol. Cell Biol. 17:2014-2019). cDNA encoding CGDD is subcloned into the pLNCX retroviral

5   vector to enable expression of antisense CGDD RNA. The resulting construct is transfected into the ecotropic BOSC23 virus-packaging cell line. Virus contained in the BOSC23 culture supernatant is used to infect the amphotropic CAK8 virus-packaging cell line. Virus contained in the CAK8 culture supernatant is used to infect normal human fibroblast (Hs68) cells. Infected cells are assessed for the following quantifiable properties characteristic of transformed cells: growth in culture to high density

10 associated with loss of contact inhibition, growth in suspension or in soft agar, formation of colonies or foci, lowered serum requirements, and ability to induce tumors when injected into immunodeficient mice. The activity of CGDD is proportional to the extent of transformation of Hs68 cells.

Alternatively, CGDD can be expressed in a mammalian cell line by transforming the cells with a eukaryotic expression vector encoding CGDD. Eukaryotic expression vectors are

15 commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. To assay the cellular localization of CGDD, cells are fractionated as described by Jiang, H.P. et al. (1992; Proc. Natl. Acad. Sci. 89:7856-7860). Briefly, cells pelleted by low-speed centrifugation are resuspended in buffer (10 mM TRIS-HCl, pH 7.4/ 10 mM NaCl/ 3 mM $MgCl_2$/ 5 mM EDTA with 10 ug/ml aprotinin, 10 ug/ml leupeptin, 10 ug/ml pepstatin A, 0.2 mM

20 phenylmethylsulfonyl fluoride) and homogenized. The homogenate is centrifuged at 600 x g for 5 minutes. The particulate and cytosol fractions are separated by ultracentrifugation of the supernatant at 100,000 x g for 60 minutes. The nuclear fraction is obtained by resuspending the 600 x g pellet in sucrose solution (0.25 M sucrose/ 10 mM TRIS-HCl, pH 7.4/ 2 mM $MgCl_2$) and recentrifuged at 600 x g. Equal amounts of protein from each fraction are applied to an SDS/10% polyacrylamide gel and

25 blotted onto membranes. Western blot analysis is performed using CGDD anti-serum. The localization of CGDD is assessed by the intensity of the corresponding band in the nuclear fraction relative to the intensity in the other fractions. Alternatively, the presence of CGDD in cellular fractions is examined by fluorescence microscopy using a fluorescent antibody specific for CGDD.

Alternatively, CGDD activity may be demonstrated as the ability to interact with its

30 associated Ras superfamily protein, in an *in vitro* binding assay. The candidate Ras superfamily proteins are expressed as fusion proteins with glutathione S-transferase (GST), and purified by affinity chromatography on glutathione-Sepharose. The Ras superfamily proteins are loaded with GDP by incubating 20 mM Tris buffer, pH 8.0, containing 100 mM NaCl, 2 mM EDTA, 5 mM MgCl2, 0.2 mM DTT, 100 µM AMP-PNP and 10 µM GDP at 30°C for 20 minutes. CGDD is

expressed as a FLAG fusion protein in a baculovirus system. Extracts of these baculovirus cells containing CGDD-FLAG fusion proteins are precleared with GST beads, then incubated with GST-Ras superfamily fusion proteins. The complexes formed are precipitated by glutathione-Sepharose and separated by SDS-polyacrylamide gel electrophoresis. The separated proteins are blotted onto

5   nitrocellulose membranes and probed with commercially available anti-FLAG antibodies. CGDD activity is proportional to the amount of CGDD-FLAG fusion protein detected in the complex.

Alternatively, as demonstrated by Li and Cohen (Li, L. and S.N. Cohen (1995) Cell 85:319-329), the ability of CGDD to suppress tumorigenesis can be measured by designing an antisense sequence to the 5' end of the gene and transfecting NIH 3T3 cells with a vector transcribing this

10   sequence. The suppression of the endogenous gene will allow transformed fibroblasts to produce clumps of cells capable of forming metastatic tumors when introduced into nude mice.

Alternatively, an assay for CGDD activity measures the effect of injected CGDD on the degradation of maternal transcripts. Procedures for oocyte collection from Swiss albino mice, injection, and culture are as described in Stutz et al., (*supra*). A decrease in the degradation of

15   maternal RNAs as compared to control oocytes is indicative of CGDD activity. In the alternative, CGDD activity is measured as the ability of purified CGDD to bind to RNAse as measured by the assays described in Example XVII.

Alternatively, an assay for CGDD activity measures syncytium formation in COS cells transfected with an CGDD expression plasmid, using the two-component fusion assay described in Mi

20   (*supra*). This assay takes advantage of the fact that human interleukin 12 (IL-12) is a heterodimer comprising subunits with molecular weights of 35 kD (p35) and 40 kD (p40). COS cells transfected with expression plasmids carrying the gene for p35 are mixed with COS cells cotransfected with expression plasmids carrying the genes for p40 and CGDD. The level of IL-12 activity in the resulting conditioned medium corresponds to the activity of CGDD in this assay. Syncytium

25   formation may also be measured by light microscopy (Mi et al., *supra*).

An alternative assay for CGDD activity measures cell proliferation as the amount of newly initiated DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding CGDD is transfected into quiescent 3T3 cultured cells using methods well known in the art. The transiently transfected cells are then incubated in the presence of [^{3}H]thymidine or a radioactive DNA

30   precursor such as [α^{32}P]ATP. Where applicable, varying amounts of CGDD ligand are added to the transfected cells. Incorporation of [^{3}H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA and CGDD activity.

Alternatively, CGDD activity is measured by the cyclin-ubiquitin ligation assay (Townsley,

F.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2362-2367). The reaction contains in a volume of 10 μl, 40 mM Tris.HCl (pH 7.6), 5 mM Mg Cl$_2$, 0.5 mM ATP, 10 mM phosphocreatine, 50 μg of creatine phosphokinase/ml, 1 mg reduced carboxymethylated bovine serum albumin/ml, 50 μM ubiquitin, 1 μM ubiquitin aldehyde, 1-2 pmol ^{125}I-labeled cyclin B, 1 pmol E1, 1 μM okadaic acid, 10

5   μg of protein of M-phase fraction·1A (containing active E3-C and essentially free of E2-C), and varying amounts of CGDD. The reaction is incubated at 18 °C for 60 minutes. Samples are then separated by electrophoresis on an SDS polyacrylamide gel. The amount of ^{125}I- cyclin-ubiquitin formed is quantified by PHOSPHORIMAGER analysis. The amount of cyclin-ubiquitin formation is proportional to the activity of CGDD in the reaction.

10   Alternatively, an assay for CGDD activity uses radiolabeled nucleotides, such as [α^{32}P]ATP, to measure either the incorporation of radiolabel into DNA during DNA synthesis, or fragmentation of DNA that accompanies apoptosis. Mammalian cells are transfected with plasmid containing cDNA encoding CGDD by methods well known in the art. Cells are then incubated with radiolabeled nucleotide for various lengths of time. Chromosomal DNA is collected, and radioactivity is detected

15   using a scintillation counter. Incorporation of radiolabel into chromosomal DNA is proportional to the degree of stimulation of the cell cycle. To determine if CGDD promotes apoptosis, chromosomal DNA is collected as above, and analyzed using polyacrylamide gel electrophoresis, by methods well known in the art. Fragmentation of DNA is quantified by comparison to untransfected control cells, and is proportional to the apoptotic activity of CGDD.

20   Alternatively, cyclophilin activity of CGDD is measured using a chymotrypsin-coupled assay to measure the rate of cis to trans interconversion (Fischer, G. et al. (1984) Biomed. Biochim. Acta 43:1101-1111). The chymotrypsin is used to estimate the trans-substrate cleavage activity at Xaa-Pro peptide bonds, wherein the rate constant for the cis to trans isomerization can be obtained by measuring the rate constant of the substrate hydrolysis at the slow phase. Samples are incubated in

25   the presence or absence of the immunosuppressant drugs CsA or FK506, reactions initiated by addition of chymotrypsin, and the fluorescent reaction measured. The enzymatic rate constant is calculated from the equation $k_{app} = k_{H20} + k_{enz}$, wherein first order kinetics are displayed, and where one unit of PPIase activity is defined as k_{enz} (s^{-1}).

          Alternatively, cyclophilin activity of CGDD is monitored by a quantitative immunoassay that

30   measures its affinity for stereospecific binding to the immunosuppressant drug cyclosporin (Quesniaux, V.F. et al. (1987) Eur. J. Immunol. 17:1359-1365). In this assay, the cyclophilin-cyclosporin complex is coated on a solid phase, with binding detected using anti-cyclophilin rabbit antiserum enhanced by an antiglobulin-enzyme conjugate.

          Alternatively, activity of CGDD is monitored by a binding assay developed to measure the

non-covalent binding between FKBPs and immunosuppressant drugs in the gas phase using electrospray ionization mass spectrometry (Trepanier, D.J. et al. (1999) Ther. Drug Monit. 21:274-280). In electrospray ionization, ions are generated by creating a fine spray of highly charged droplets in the presence of a strong electric field; as the droplet decreases in size, the charge density on the

5   surface increases. Ions are electrostatically directed into a mass analyzer, where ions of opposite charge are generated in spatially separate sources and then swept into capillary inlets where the flows are merged and where reactions occur. By comparing the charge states of bound versus unbound CGDD/immunosuppressive drug complexes, relative binding affinities can be established and correlated with *in vitro* binding and immunosuppressive activity.

10   **XIX. CGDD Secretion Assay**

A high throughput assay may be used to identify polypeptides that are secreted in eukaryotic cells. In an example of such an assay, polypeptide expression libraries are constructed by fusing 5'-biased cDNAs to the 5'-end of a leaderless β-lactamase gene. β-lactamase is a convenient genetic reporter as it provides a high signal-to-noise ratio against low endogenous background activity and

15   retains activity upon fusion to other proteins. A dual promoter system allows the expression of β-lactamase fusion polypeptides in bacteria or eukaryotic cells, using the *lac* or CMV promoter, respectively.

Libraries are first transformed into bacteria, *e.g.*, *E. coli*, to identify library members that encode fusion polypeptides capable of being secreted in a prokaryotic system. Mammalian signal

20   sequences direct the translocation of β-lactamase fusion polypeptides into the periplasm of bacteria where it confers antibiotic resistance to carbenicillin. Carbenicillin-selected bacteria are isolated on solid media, individual clones are grown in liquid media, and the resulting cultures are used to isolate library member plasmid DNA.

Mammalian cells, *e.g.*, 293 cells, are seeded into 96-well tissue culture plates at a density of

25   about 40,000 cells/well in 100 μl phenol red-free DME supplemented with 10% fetal bovine serum (FBS) (Life Technologies, Rockville, MD). The following day, purified plasmid DNAs isolated from carbenicillin-resistant bacteria are diluted with 15 μl OPTI-MEM I medium (Life Technologies) to a volume of 25 μl for each well of cells to be transfected. In separate plates, 1 μl LF2000 Reagent (Life Technologies) is diluted into 25 μl/well OPTI-MEM I. The 25 μl diluted LF2000 Reagent is

30   then combined with the 25 μl diluted DNA, mixed briefly, and incubated for 20 minutes at room temperature. The resulting DNA-LF2000 reagent complexes are then added directly to each well of 293 cells. Cells are also transfected with appropriate control plasmids expressing either wild-type β-lactamase, leaderless β-lactamase, or, for example, CD4-fused leaderless β-lactamase. 24 hrs following transfection, about 90 μl of cell culture media are assayed at 37°C with 100 μM Nitrocefin

(Calbiochem, San Diego, CA) and 0.5 mM oleic acid (Sigma Corp. St. Louis, MO) in 10 mM phosphate buffer (pH 7.0). Nitrocefin is a substrate for β-lactamase that undergoes a noticeable color change from yellow to red upon hydrolysis. β-lactamase activity is monitored over 20 min in a microtiter plate reader at 486 nm. Increased color absorption at 486 nm corresponds to secretion of a

5   β-lactamase fusion polypeptide in the transfected cell media, resulting from the presence of a eukaryotic signal sequence in the fusion polypeptide. Polynucleotide sequence analysis of the corresponding library member plasmid DNA is then used to identify the signal sequence-encoding cDNA. (Described in U.S. Patent application 09/803,317, filed March 9, 2001.)

For example, SEQ ID NO:14 was shown to be a secreted protein using this assay.

10   Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions.

15   Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of

20   embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
261510	1	261510CD1	28	261510CB1	90110645CA2
7498674	2	7498674CD1	29	7498674CB1	90175915CA2, 90175931CA2
1629617	3	1629617CD1	30	1629617CB1	
2369279	4	2369279CD1	31	2369279CB1	
3167506	5	3167506CD1	32	3167506CB1	
3075937	6	3075937CD1	33	3075937CB1	
5176268	7	5176268CD1	34	5176268CB1	90110594CA2
72830854	8	72830854CD1	35	72830854CB1	
3632052	9	3632052CD1	36	3632052CB1	
7493817	10	7493817CD1	37	7493817CB1	
6715627	11	6715627CD1	38	6715627CB1	
7727886	12	7727886CD1	39	7727886CB1	
914113	13	914113CD1	40	914113CB1	
1953711	14	1953711CD1	41	1953711CB1	
1595275	15	1595275CD1	42	1595275CB1	
72332548	16	72332548CD1	43	72332548CB1	
7322834	17	7322834CD1	44	7322834CB1	
1925714	18	1925714CD1	45	1925714CB1	
6803363	19	6803363CD1	46	6803363CB1	90192087CA2
7070580	20	7070580CD1	47	7070580CB1	90136522CA2
7500176	21	7500176CD1	48	7500176CB1	
7500506	22	7500506CD1	49	7500506CB1	90052311CA2
7500639	23	7500639CD1	50	7500639CB1	1355056CA2, 90133101CA2, 90133117CA2, 90133118CA2, 90133225CA2, 90133233CA2

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7506167	24	7506167CD1	51	7506167CB1	
90081189	25	90081189CD1	52	90081189CB1	90081189CA2, 90081265CA2, 90129619CA2
7510095	26	7510095CD1	53	7510095CB1	
7510096	27	7510096CD1	54	7510096CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
2	7498674CD1	g4091819	4.30E-161	[Homo sapiens] leucine-rich glioma-inactivated protein precursor (Chernova, O.B. et al. (1998) Oncogene 17:2873-2881)
3	1629617CD1	g9965418	1.50E-190	[Mus musculus] iroquois-class homeobox protein IRX2 Christoffels, V.M. et al. Patterning the embryonic heart: identification of five mouse iroquois homeobox genes in the developing heart Dev. Biol. 224, 263-274 (2000)
4	2369279CD1	g4581563	1.90E-64	[Homo sapiens] tesmin Sugihara,T. et al. A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation Genomics 57, 130-136 (1999)
5	3167506CD1	g12697318	2.60E-177	[Homo sapiens] PBX4 protein Wagner,K. et al. Pbx4, a new Pbx family member on mouse chromosome 8, is expressed during spermatogenesis Mech. Dev. 103, 127-131 (2001)
6	3075937CD1	g2072425	0.0	[Homo sapiens] non-lens beta gamma-crystallin like protein Ray, M.E. et al. (1997) AIM1, a novel non-lens member of the betagamma-crystallin superfamily, is associated with the control of tumorigenicity in human malignant melanoma. Proc. Natl. Acad. Sci. U.S.A. 94:3229-3234
7	5176268CD1	g14530665	1.10E-38	[Homo sapiens] (AB052907) ALCAN-beta
		g19525540	1.00E-148	[fl][Homo sapiens] lymphocyte effector toxicity activation ligand
8	72830854CD1	g3170204	9.90E-162	[Homo sapiens] antigen NY-CO-43 Scanlan, M.J. et al. (1998) Characterization of human colon cancer antigens recognized by autologous antibodies. Int. J. Cancer 76:652-658

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
9	3632052CD1	g4105619	6.40E-218	[Mus musculus] SPAF (spermatogenesis associated factor, AAA family) Liu, Y. et al. (2000) Oncogene 19:1579-1588 SPAF, a new AAA-protein specific to early spermatogenesis and malignant conversion.
10	7493817CD1	g6062874	5.20E-236	[Homo sapiens] candidate tumor suppressor protein DICE1 Wieland, I. et al. (1999) Oncogene 18:4530-4537 Isolation of DICE1: a gene frequently affected by LOH and downregulated in lung carcinomas
11	6715627CD1	g2231999	0.0	[Homo sapiens] MEA6 Heckel,D. et al. (1997) Hum. Mol. Genet. 6:2031-2041 cDNA cloning and chromosomal mapping of a predicted coiled-coil proline-rich protein immunogenic in meningioma patients
12	7727886CD1	g1220353	2.40E-150	[Homo sapiens] paraneoplastic antigen Fathallah-Shaykh, H. et al. (1991) Proc. Natl. Acad. Sci. USA.88:3451-3454 Cloning of a leucine-zipper protein recognized by the sera of patients with antibody-associated paraneoplastic cerebellar degeneration.
13	914113CD1	g14861177	8.10E-73	[Danio rerio] (AF222996) nocA-like Zn-finger protein (Sagerstrom,C.G. et al. (2001) Dev. Dyn. 220 (4), 402-408)
14	1953711CD1	g1177322	0.0	[Rattus norvegicus] CPG2 protein (Nedivi,E. et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93 (5), 2048-2053)
15	1595275CD1	g1297319	7.70E-44	[Homo sapiens] DOC1 Mok,S.C. et al. (1994) Gynecol. Oncol. 52:247-252
16	72332548CD1	g5730480	3.00E-91	[Homo sapiens] candidate tumor suppressor p33 ING1 homolog
17	7322834CD1	g150077051	0.0	[Mus musculus] (AF288137) metastasis associated protein 1 Simpson,A. et al. (2001) Gene 273:29-39
18	1925714CD1	g12659140	8.00E-256	[Mus musculus] mage-e1
19	6803363CD1	g3253213	2.80E-43	[Homo sapiens] glioma amplified on chromosome 1 protein Almeida,A. et al. (1998) Oncogene 16:2997-3002
20	7070580CD1	g3041877	6.50E-176	[Homo sapiens] IB3089A Habuchi,T. et al. (1998)Genomics 48: 277-288

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
21	7500176CD1	g1322234	0.0	[Homo sapiens] OS-9 precurosor Su,Y.A. et al. (1996) Mol. Carcinog. 15:270-275
22	7500506CD1	g13097582	4.60E-241	[Homo sapiens] (BC003512) mesothelin
23	7500639CD1	g14039857	1.20E-23	[Homo sapiens] testes development-related NYD-SP22
24	7506167CD1	g1916672	0.0	[Homo sapiens] meningioma-expressed antigen 11 Heckel,D. et al. (1997) cDNA cloning and chromosomal mapping of a predicted coiled-coil proline-rich protein immunogenic in meningioma patients. Hum. Mol. Genet. 6:2031-2041
		343086|MGEA6	1.5E-244	[Homo sapiens][Activator; Small molecule-binding protein] Meningioma expressed antigen 6, a tumor antigen with an acidic activation domain, a proline-rich region and two coiled-coil domains Heckel,D. et al. (1997) cDNA cloning and chromosomal mapping of a predicted coiled-coil proline-rich protein immunogenic in meningioma patients. Hum. Mol. Genet. 6:2031-2041
		6350|USO1	1.2E-13	[Saccharomyces cerevisiae][Docking protein][Golgi] Coiled-coil protein with similarity to mammalian integrin, required for protein transport from ER to Golgi Nakajima, H.et al. (1991) A cytoskeleton-related gene, uso1, is required for intracellular protein transport in Saccharomyces cerevisiae. J. Cell Biol. 113:245-260
25	90081189CD1	g1322234	0.0	[Homo sapiens] OS-9 precurosor Su, Y. A. et al. Complete sequence analysis of a gene (OS-9) ubiquitously expressed in human tissues and amplified in sarcomas. Mol Carcinog 15, 270-5

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	90081189CD1	428204|OS-9	1.3E-291	[Homo sapiens] Amplified in osteosarcoma 9, a putative soluble acidic protein that may positively regulate cell growth and proliferation; co-amplification of the corresponding gene with cyclin-dependent kinase 4 (CDK4) is associated with sarcomas. Su, Y. A. et al. (1996) Complete sequence analysis of a gene (OS-9) ubiquitously expressed in human tissues and amplified in sarcomas. Mol. Carcinog. 15:270-275.
26	7510095CD1	g1322234	0.0	[Homo sapiens] OS-9 precurosor Su, Y. A. et al. (supra)
		742638|CL25084	6.8E-19	[Homo sapiens] Protein of unknown function, has a region of low similarity to a region of OS-9, which is an acidic soluble protein that is associated with human sarcomas.
27	7510096CD1	g1322234	8.50E-21	[Homo sapiens] OS-9 precurosor. Su, Y. A. et al. (supra)
		428204|OS-9	7.0E-22	[Homo sapiens] Amplified in osteosarcoma 9, a putative soluble acidic protein that may positively regulate cell growth and proliferation; coamplification of the corresponding gene with cyclin-dependent kinase 4 (CDK4) is associated with sarcomas. Su, Y. A. et al. (supra) Kimura, Y. et al. Genomic organization of the OS-9 gene amplified in human sarcomas. J Biochem (Tokyo) 122, 1190-5 (1997) Nakayama, T. et al. Ca2(+)-dependent interaction of N-copine, a member of the two C2 domain protein family, with OS-9, the product of a gene frequently amplified in osteosarcoma. FEBS Lett 453, 77-80 (1999)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	261510CD1	168	S25 S70 S90 T120		PROTEIN EREBP4 HOMOLOG CODED FOR BY C ELEGANS CDNA YK103H12.5 PD022373: V42-L131, L6-S17	BLAST_PRODOM
2	7498674CD1	548	S29 S97 S348 S531 T131 T312 T377 T407	N189 N311	signal_cleavage: M1-A30	SPSCAN
					Signal Peptide: M1-A30, R7-A30	HMMER
					Leucine Rich Repeat: S137-D160, L89-S112, H113-K136, E65-P88	HMMER_PFAM
					Leucine Rich Repeat C-terminal domain: N170-P212, L213-I219	HMMER_PFAM
					Lipocalin signature: D331-A344	MOTIFS
3	1629617CD1	378	S93 S106 S121 S133 S143 S159 S238 T35 T50 T120 T196	N27 N91 N298	Homeobox domain: R44-K81	HMMER_PFAM
					'Homeobox' domain signature and profile: T33-T104	PROFILESCAN
					'Homeobox' domain protein BL00027: W39-K81	BLIMPS_BLOCKS
					Homeobox signature PR00024: N46-L57, T61-W71, W71-K80	BLIMPS_PRINTS
					IROQUOIS CLASS HOMEODOMAIN PROTEIN IRX2 HOMEOBOX DNA BINDING NUCLEAR PD086674: S232-N304	BLAST_PRODOM
					HOMEOBOX DM00009|P54269|222-288: A23-K84	BLAST_DOMO
					'Homeobox' domain signature: L57-K80	MOTIFS
4	2369279CD1	528	S370 S375 T65	N77 N228 N271 N305 N329	DNA BINDING DOMAIN PROTEIN-LIKE CXC FACTOR TSO1 CONTAINING ACTIVATING JC8.6B PD042633: K313-N410	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	3167506CD1	374	S33 S67 S140 S170 S195 S318 T202 T283 T286 T325 Y282	N216 N369	Homeobox domain: R211-K270	HMMER_PFAM
					'Homeobox' domain signature and profile: T222-E292	PROFILESCAN
					'Homeobox' domain protein BL00027: Y228-K270	BLIMPS_BLOCKS
					Homeobox signature PR00024: N235-L246, G250-W260, W260-K269	BLIMPS_PRINTS
					HOMEOBOX PROTEIN DNA BINDING NUCLEAR TRANSCRIPTION REGULATION PREB CELL LEUKEMIA ACTIVATOR ALTERNATIVE PD007515: D18-A210	BLAST_PRODOM
					HOMEOBOX; EXTRADENTICLE; DM08777 |A54863|54-177: T30-H155 |P40427|56-182: Q32-H155	BLAST_DOMO
					HOMEOBOX DM00009 |P40425|240-306: R206-G273 |P40424|229-295: R206-G273	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop): A279-T286	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	3075937CD1	1723	S6 S10 S16 S61 S92 S151 S209 S274 S284 S299 S348 S456 S558 S626 S689 S693 S695 S759 S785 S827 S900 S915 S920 S979 S989 S990 S1002 S1028 S1033 S1104 S1112 S1123 S1139 S1158	N52 N422 N681 N867 N907 N1462	Beta/Gamma crystallin: K1219-I1312, T1503-F1583, H1320-I1403, Q1416-I1495, H1124-L1205	HMMER_PFAM
			S1190 S1201 S1265 S1308 S1325 S1334 S1386 S1431 S1437 S1444 S1447 S1491 S1552 S1561 S1579 S1672		BETA-CRYSTALLIN DM00094 [P07318]145-233: V1263-R1310 [P11842]21-111: K1219-I1312 [P19141]10-199: V1263-I1312	BLAST_DOMO
			T105 T147 T228 T246 T339 T346 T426 T432 T450 T530 T931 T1008 T1129 T1253 T1385 T1488 T1695		Crystallins beta and gamma 'Greek key' motif signature: W1064-L1079, W1166-L1181, V1220-L1235, W1273-L1288, W1364-L1379, W1455-L1470, I1504-L1519	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	5176268CD1	255	S5 S185 S222 T38 T86 T93 T142 T170 T255	N36 N82 N154 N212	signal_cleavage: M1-G29	SPSCAN
					Signal Peptide: S8-G30, M1-M27, M1-G30, M1-S32	HMMER
					Cytosolic domain: M1-R12, Q249-T255 Transmembrane domain: L13-F35, R226-W248 Non-cytosolic domain: N36-D225	TMHMMER
					RETINOIC ACID EARLY TRANSCRIPT CELL SURFACE PROTEIN MHC ALPHA GAMMA PD023817: L18-P224	BLAST_PRODOM
					CLASS I HISTOCOMPATIBILITY ANTIGEN DM00083|P13753|9-210: P10-L197	BLAST_DOMO
8	72830854CD1	1278	S318 S347 S352 S527 S601 S637 S702 S706 S724 S786 S822 S872 S901 S925 S937 S953 S968 S982 S1055 S1059 S1072 S1123 S1274 T348 T388 T418 T438 T460 T481 T578 T721 T747 T862 T926 T1014 T1116 T1122 T1181 T1194 T1213 Y1096	N1057 N1173	Signal Peptide: M1-P25, M1-P28, M1-A26, M44-A69	HMMER
					HUMAN COLON CANCER ANTIGEN NYCO43 PD133698: M809-Q965	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8					KINASE WNK1 KIAA0344 ANTIGEN NYCO43 PD041299: T966-L1154	BLAST_PRODOM
					LUMICAN; FIBROMODULIN; T-CELL; CD2; DM02006|A40678|300-393: P113-A189	BLAST_DOMO
					MICROSATELLITE BAT2; DM05517 |P48634|1-1860: A17-K300 |S37671|1-1870: K33-I277	BLAST_DOMO
					Leucine zipper pattern: L13-L34	MOTIFS
					Cell attachment sequence: R762-D764	MOTIFS
9	3632052CD1	541	S2 S3 S26 S112 S171 S173 S185 S190 S199 S237 S258 T210 T224 T267 T292 T298 T361 T406 T442 T462 T508 Y61 Y287	N11 N16 N296	Signal_cleavage: M1-A30	SPSCAN
					Non-cytosolic domain:M1-W541	TMHMMER
					ATPase family associated with various cellular activities (chaperone): G352-L540	HMMER_PFAM
					AAA-protein family proteins BL00674: Y315-P335, P350-A371, G383-R425, T442-N488, G521-L540	BLIMPS_BLOCKS
					PROTEIN ATPBINDING PROTEASE SUBUNIT HOMOLOG REPEAT CELL DIVISION ATPDEPENDENT NUCLEAR PD000092: G352-Q497	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9					AAA-PROTEIN FAMILY DM00024 |Q07590|207-365: K312-R473 |P54609|201-360: V313-R473 |P23787|198-357: K312-R473 |P03974|198-357: K312-R473	BLAST_DOMO
					AAA-protein family signature: V456-R474	MOTIFS
					ATP/GTP-binding site motif A (P-loop): G357-T364	MOTIFS
10	7493817CD1	553	S23 S45 S94 S106 S253 S314 S359 S464 S471 T54 T127 T181 T212 T238 T467 Y367	N492	signal_cleavage: M1-Y24	SPSCAN
					Non-cytosolic domain:M1-V553	TMHMMER
					TRANSMEMBRANE CODED FOR BY C ELEGANS CDNA YK25A9.5 NOTCH2-LIKE EGF PD043679: E121-R507	BLAST_PRODOM
					Cell attachment sequence: R46-D48	MOTIFS
11	6715627CD1	804	S62 S65 S111 S137 S139 S160 S170 S175 S182 S221 S297 S325 S341 S415 S577 S613 S631 S657 S661 S662 T193 T245 T266 T268 T286 T346 T369 T401 T430 T489 T539 T666 Y406	N135 N323 N664	signal_cleavage: M1-R64	SPSCAN
					Signal Peptide: M30-A54	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11					Cytosolic domain:M1-Y37 Transmembrane domain:G38-W60 Non-cytosolic domain:R61-T804	TMHMMER
					Molluscan rhodopsin C-terminal tail signature PR00239: P755-Y762, Y754-R766, P755-R766	BLIMPS_PRINTS
					MEA6 PD123689: S448-G710, P767-P773 PD085638: M1-E102 PD075273: V244-K310	BLAST_PRODOM
					PROTEIN COILED COIL CHAIN MYOSIN REPEAT HEAVY ATP-BINDING FILAMENT HEPTAD PD000002: E107-E300	BLAST_PRODOM
					Myosin-like protein DM07884|Q02455|35-1728: R74-I446	BLAST_DOMO
					TRICHOHYALIN DM03839|P37709|632-1103: E87-R444	BLAST_DOMO
					Leucine zipper pattern: L127-L148	MOTIFS
12	7727886CD1	465	S11 S17 S54 S167 S179 S228 S316 S344 S397 S432 S463 T37 T55 T72 T90 T154 T388 T450 Y220 Y360	N337 N455	LEUKOCYTE DNA BINDING RECEPTOR CDR2 PARANEOPLASTIC CEREBELLAR DEGENERATION ASSOCIATED ANTIGEN PD024533: W18-K460	BLAST_PRODOM
					Non-cytosolic:M1-K465	TMHMMER
					Cell attachment sequence: R292-D294	MOTIFS
					Leucine zipper pattern: L115-L136, L181-L202	MOTIFS
13	914113CD1	651	S12 S56 S106 S135 S178 S187 S211 S604 T85 T227 T536 T553	N43	Type I antifreeze protein signature PR00308: C442-C456, T488-P497	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13					LORICRIN; DM07383|P18165|166-480: S135-G340, K14-G55, C442-T477 |P23490|22-315: S139-V342, H15-F60	BLAST_DOMO
					PHASEOLUS GLYCINE-RICH CELL WALL PROTEIN 1.8 DM07973 |P09789|1-383: G145-A343, P5-G55 |P27483|1-337: G145-G340, R11-G55, N144-G157	BLAST_DOMO
14	1953711CD1	1443	S153 S168 S206 S234 S350 S494 S535 S545 S576	N110 N287 N588 N596 N637 N1143	Calponin homology (CH) domain: G178-P283, I27-Q134	HMMER_PFAM
			S603 S606 S613 S732 S777 S826 S935 S968 S1034 S1054 S1066 S1141		Actinin-type actin-binding domain proteins BL00019: Q29-S39, D56-E78, K103-L138	BLIMPS_BLOCKS
			S1196 S1213 S1222 S1244 S1286 S1319 S1361 S1370 S1377 S1386 S1422 T3 T173 T196 T229		Actinin-type actin-binding domain signatures: R9-V61, K103-S147	PROFILESCAN
			T244 T393 T670 T839 T870 T964 T1040 T1104 T1115 T1130 T1175 T1282 T1338 T1347 T1433 Y1116		CPG2 PROTEIN PD148409: M501-I1440	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14					PROTEIN ACTIN-BINDING REPEAT CALCIUM BINDING FIMBRIN PHOSPHORYLATION ALPHA-ACTININ MULTIGENE FAMILY CYTOSKELETON PD000807: Q25-I130	BLAST_PRODOM
					ALPHA-ACTININ ACTININ-BINDING DOMAIN DM00325 |P30427|179-405: D23-L138, A180-P283 |A44159|48-277: L21-T139, K175-S234 |P11277|48-260: L21-V159, N179-P266 |P46939|25-254: D23-S150, Q177-P283	BLAST_DOMO
					Actinin-type actin-binding domain signature 1: Q29-N38	MOTIFS
15	1595275CD1	1087	S6 S12 S24 S108 S130 S140 S152 S153 S167 S181 S195 S233 S255 S275 S307 S338 S366 S417	N58 N101 N146 N231 N269 N273 N413 N548 N675 N715 N1019	Aldehyde ferredoxin oxid PF01314: D850-P886, T300-I343, F407-M437, E99-K131	BLIMPS_PFAM
			S459 S460 S569 S570 S573 S574 S608 S622 S639 S652 S659 S703 S717 S720 S740 S745 S772		PROTEIN COILED COIL CHAIN MYOSIN REPEAT HEAVY ATP-BINDING FILAMENT HEPTAD PD000002: E76-K316	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Glycosylation Sites	Potential Phosphorylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15				S805 S840 S926 S932 S956 S961 S981 S989 S990 S1023 S1056 T148 T215 T389 T427 T448 T491	CHROMOSOME PROTEIN COILED COIL HEPTAD REPEAT PATTERN ATP-BINDING I MAJOR PD075049: M92-E361	BLAST_PRODOM
				T507 T510 T533 T670 T776 T898 T919 T953 T975 T998 T1004	TRICHOHYALIN DM03839|P37709|632-1103: L16-E412	BLAST_DOMO
					INTERMEDIATE FILAMENTS DM00061|P23731|37-456: L27-L353	BLAST_DOMO
					MYOSIN-LIKE PROTEIN MLP1 DM07884|Q02455|35-1728: L16-S579	BLAST_DOMO
					TRICHOHYALIN DM03839|P22793|921-1475: R14-Q420	BLAST_DOMO
					Leucine zipper pattern: L16-L37, L23-L44, L158-L179, L165-L186	MOTIFS
					Mitochondrial energy transfer proteins signature: P640-I649	MOTIFS
16	72332548CD1	240		S118 S140 S148 S195 T52 T147 T152 T220	PHD-finger: Y188-E235	HMMER_PFAM
					PHD-finger PF00628: P205-L219	BLIMPS_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	7322834CD1	686	S119 S122 S175 S215 S272 S305 S443 S644 T189 T316 T417 T458 T496 T636 T648 Y13 Y380	N16 N31 N115 N361 N427	signal_cleavage: M1-A63	SPSCAN
					BAH domain: N4-K147	HMMER_PFAM
					ELM2 domain: G148-A209	HMMER_PFAM
					GATA zinc finger: A372-K414	HMMER_PFAM
					Myb-like DNA-binding domain: M268-K314	HMMER_PFAM
					METASTASIS-ASSOCIATED PROTEIN MTA1 PD040130: M415-L667 PD023997 (T27C4.4): E65-L177 PD011563 (SIMILAR MTA1 T27C4.4 KIAA0458 C04A2.2 CHROMOSOME II): D231-R324 PD024227: L325-G371	BLAST_PRODOM
18	1925714CD1	1595	S60 S111 S159 S198 S205 S227 S238 S288 S345 S528 S564 S646 S662 S782 S1073 S1083 S1257	N7 N123 N383 N660 N775 N1115 N1447	MAGE family: R652-Q1251, I1313-Q1505	HMMER_PFAM
			S1297 S1370 S1377 S1381 S1399 S1518 S1524 T176 T181 T215 T279 T312 T321 T327 T354		DNA RNA-DIRECTED POLYMERASE PUTATIVE P150 PROTEIN TRANSCRIPTASE REVERSE L1 SEQUENCE PD003002: T257-W371	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18			T456 T459 T467 T481 T504 T509 T517 T533 T548 T709 T721 T733 T745 T776 T781 T793 T805		PUTATIVE P150 DNA PROTEIN RNA-DIRECTED POLYMERASE TRANSCRIPTASE REVERSE L1 SEQUENCE PD002970: H393-G472 PD002839: M473-I552	BLAST_PRODOM
			T817 T829 T841 T865 T877 T889 T901 T937 T949 T973 T985 T1009 T1021		ANTIGEN MELANOMA-ASSOCIATED MULTIGENE FAMILY PROTEIN TUMOR RELATED POLYMORPHISM MAGE4 MAGEB1 PD003141: Q1145-G1249	BLAST_PRODOM
			T1033 T1116 T1122 T1301 T1513 T1547 Y23 Y1158		TRANSCRIPTASE; REVERSE; II; ORF2; DM01354|P08547|558-973: Q232-K458 DM01354|I38588|559-974: M244-K458 DM01354|TU0033|48-463: Q232-K458 DM01354|S23650|1-411: M244-K458	BLAST_DOMO
					Leucine zipper pattern: L212-L233	MOTIFS
19	6803363CD1	592	S151 S166 S212 S242 S289 S331 S568 T41 T46	N127 N185 N247 N257 N276 N324 N488 N512 N562	Signal Peptide: M1-P20, M1-G24	HMMER
			T123 T129 T145 T332 T346 T369 T381 T420 T496 T506 Y478		Leucine Rich Repeat: S151-L174 N247-A270	HMMER_PFAM
					Leucine Rich Repeat: H271-V294, A79-P102, R103-D126, A175-R198, N127-H150, Q319-N342, E55-P78, R295-R318, S199-P222	HMMER_PFAM
					Leucine rich repeat C-terminal domain: N352-R405	HMMER_PFAM
					Leucine rich repeat N-terminal domain: G24-P53	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19					Immunoglobulin domain: G422-A482	HMMER_PFAM
					Cytosolic domain: W552-I592; Transmembrane domain: V529-V551; Non-cytosolic domain: M1-L528	TMHMMER
					Leucine-rich repeat signature PR00019: L80-V93, L101-L114	BLIMPS_PRINTS
					GLYCOPROTEIN HORMONE RECEPTOR DM0297|P35376|14-202: G24-L181	BLAST_DOMO
					Cell attachment sequence: R386-D388	MOTIFS
20	7070580CD1	678	S34 S130 S179 S369 S418 S435 S440 S451 S475 S491 S504 S547 S551 S657	N80 N249 N368 N474 N521 N553	IB3089A PD147692: T40-C677	BLAST_PRODOM
			T47 T103 T197 T281 T330 T448 T511 T532 T673		FKBP-type peptidyl-prolyl cis-trans isomerase signature 1: L319-S335	MOTIFS
21	7500176CD1	652	S64 S150 S179 S361 S447 S454 S469 T158 T221 T224 T302 T376 T456 T518 T543 T596 Y93 Y120 Y268 Y632	N177	signal_cleavage: M1-T25	SPSCAN
					Signal Peptide: M1-P21; M1-A22; M1-T25; M1-G27; M1-G26; M1-P21	HMMER
21					Cytosolic domain: M1-L6 Transmembrane domain: L7-G29 Non-cytosolic domain: S30-F652	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					CheB methylesterase. PF01339: G178-E191	BLIMPS_PFAM
					PROTEIN OS9 PRECURSOR SIGNAL PD141024: L232-F652	BLAST_PRODOM
					ALTERNATIVE AMPLIFIED PRECURSOR OS-9 SIMILAR SPLICING SIGNAL OSTEOSARCOMA PD090541: P21-P99	BLAST_PRODOM
					PROTEIN F7N22.4 OS9 PRECURSOR SIGNAL PD033854: E100-P231	BLAST_PRODOM
					Cell attachment sequence: R404-D406	MOTIFS
22	7500506CD1	457	S62 S110 S115 S139 S215 S280 S427 T5 T41 T79 T169 T225 T229 T248 T370 Y195 Y312	N57 N223 N323 N350	signal_cleavage: M1-P33	SPSCAN
					Signal Peptide: M1-A38	HMMER
					Cytosolic domain: M1-P8 Transmembrane domain: L9-V31 Non-cytosolic domain: Q32-A457	TMHMMER
					PRECURSOR SIGNAL POTENTIATING FACTOR MESOTHELIN CAK1 ANTIGEN PREPROMEGAKARYOCYTE MEGAKARYOCYTE PD025470: L77-A457	BLAST_PRODOM
23	7500639CD1	127	S67 S103 T44 T86		Hemopexin domain signature: F152-L166	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
24	7506167CD1	763	S29 S62 S65 S110 S111 S138 S140 S161 S176 S183 S222 S299 S327 S343 S417 S536 S572 S590 S616 S620 S621 T197 T247 T287 T288 T348 T371 T403 T432 T491 T625 Y37 Y408	N136 N325 N623	signal_cleavage: M1-R64	SPSCAN
					Cytosolic domain: M1-Y37 Transmembrane domain: G38-W60 Non-cytosolic domain: R61-T763	TMHMMER
					ATP synthase B/B' CF(0). PF00430: V45-E87	BLIMPS_PFAM
					Meningioma-expressed antigen 6/11 (MEA6) PD123689: G515-G669, S450-P682, P726-P732	BLAST_PRODOM
					Meningioma-expressed antigen 6/11 (MEA6) PD085638: M1-E102	BLAST_PRODOM
					Meningioma-expressed antigen 6/11 (MEA6) PD075273: I246-K312	BLAST_PRODOM
					Meningioma-expressed antigen 6/11 (MEA6) PD046284: H734-T763	BLAST_PRODOM
					MYOSIN-LIKE PROTEIN MLP1 DM07884|Q02455|35-1728: L82-E501, R74-I448	BLAST_DOMO
					Leucine zipper pattern: L128-L149	MOTIFS
25	90081189CD1	612	S64 S150 S179 S361 S447 S484 T158 T221 T224 T302 T376 T471 T533 T556 Y93 Y120 Y268 Y592	N177	signal_cleavage: M1-T25	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
25					Signal Peptide: M1-P21, M1-A22, M1-T25, M1-G27, M1-G26	HMMER
					Cytosolic domain: M1-L6 Transmembrane domain: L7-G29 Non-cytosolic domain: S30-F612	TMHMMER
					PROTEIN F7N22.4 OS9 PRECURSOR SIGNAL PD033854: E100-P231	BLAST_PRODOM
					PROTEIN OS9 PRECURSOR SIGNAL PD090541: P21-P99	BLAST_PRODOM
					PROTEIN OS9 PRECURSOR SIGNAL PD141024: L232-T533	BLAST_PRODOM
					95.1 KD PROTEIN F48E8.4 IN CHROMOSOME III PROTEIN PD148072: S55-K271	BLAST_PRODOM
					Cell attachment sequence: R404-D406	MOTIFS
26	7510095CD1	638	S64 S150 S179 S361 S447 S484 S604 T158 T221 T224 T302 T376 T471 T533 T558 Y93 Y120 Y268	N177	Signal cleavage: M1-T25	SPSCAN
					Signal Peptide: M1-P21, M1-A22, M1-T25, M1-G27, M1-G26	HMMER
					Cytosolic domain: M1-L6 Transmembrane domain: L7-G29 Non-cytosolic domain: S30-N638	TMHMMER
					PROTEIN OS9 PRECURSOR SIGNAL PD141024: L232-W608, E100-P231, P21-P9	BLAST_PRODOM
					Cell attachment sequence: R404-D406	MOTIFS
27	7510096CD1	75			Signal cleavage: M1-T25	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
27					Signal Peptide: M1-P21, M1-A22, M1-G26, M1-G27	HMMER
					Cytosolic domain: M1-L6 Transmembrane domain: L7-G29 Non-cytosolic domain: S30-W75	TMHMMER
					PROTEIN OS9 PRECURSOR SIGNAL PD090541: P21-Q54	BLAST_PRODOM

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
28/261510CB1/ 3343	1-535, 14-233, 36-275, 36-593, 36-726, 52-430, 56-349, 59-260, 67-264, 67-505, 67-538, 74-324, 108-435, 122-723, 127-595, 190-565, 219-870, 240-479, 249-843, 278-524, 278-539, 278-554, 282-925, 289-545, 332-948, 347-613, 368-945, 428-808, 518-966, 520-962, 522-966, 525-792, 525-966, 526-1086, 544-964, 554-963, 556-963, 559-966, 563-964, 568-964, 577-964, 581-962, 582-873, 594-958, 603-962, 605-967, 623-962, 632-959, 644-1260, 667-896, 716-1340, 949-1601, 954-1583, 1164-1462, 1308-1592, 1429-1634, 1429-2062, 1739-1986, 1743-1976, 1776-2053, 1815-2050, 1821-2109, 1822-2046, 1822-2419, 1837-2032, 1837-2403, 1952-2309, 1991-2281, 2021-2279, 2252-2458, 2266-2514, 2302-2564,
	2302-2593, 2310-2654, 2333-2578, 2333-2601, 2340-2575, 2364-2543, 2367-2622, 2387-2655, 2404-2662, 2418-2628, 2432-2715, 2433-2737, 2449-2634, 2449-2726, 2449-2740, 2453-2737, 2469-2664, 2469-3077, 2494-2692, 2507-2749, 2526-2775, 2526-2776, 2530-2760, 2531-3277, 2642-2842, 2667-2944, 2679-3002, 2717-3328, 2735-2977, 2735-3171, 2735-3314, 2737-3327, 2746-2998, 2749-2998, 2837-3321, 2838-3050, 2838-3205, 2851-3289, 2854-3075, 2854-3323, 2854-3343, 2898-3128, 2898-3313, 2898-3343, 2923-3337, 2983-3324, 3049-3277, 3183-3332
29/7498674CB1/ 3190	1-211, 1-477, 1-507, 1-550, 1-552, 1-556, 1-601, 1-614, 1-623, 1-634, 1-639, 1-641, 1-643, 1-657, 2-691, 10-665, 39-169, 141-799, 366-970, 712-1005, 712-1075, 712-1097, 712-1198, 845-1037, 903-1000, 925-1495, 957-1229, 957-1574, 1064-1562, 1188-1741, 1214-1746, 1238-1924, 1269-1911, 1293-1593, 1300-1601, 1318-1890, 1336-1736, 1378-1725, 1378-1736, 1410-1930, 1417-2031, 1452-1736, 1520-1771, 1577-1879, 1577-2205, 1626-1968, 1634-2149, 1688-1959, 1720-2262, 1720-2276, 1720-2297, 1720-2340, 1724-2245, 1730-2373, 1762-2042, 1774-2048, 1801-2079, 1867-2391, 1915-2196, 1917-2184,
	1920-2190, 1949-2251, 2041-2312, 2076-2327, 2076-2332, 2076-2542, 2095-2372, 2095-2556, 2185-2472, 2186-2446, 2187-2332, 2248-2503, 2248-2715, 2259-2492, 2259-2785, 2265-2507, 2279-2533, 2348-2985, 2355-2620, 2356-2830, 2356-2836, 2356-2990, 2358-2990, 2370-2663, 2378-2986, 2385-2957, 2385-2989, 2428-2995, 2449-2728, 2498-2988, 2501-2988, 2503-2985, 2513-2765, 2527-2782, 2553-2969, 2592-3005, 2593-2840, 2602-2989, 2610-3190, 2611-2915, 2612-2990, 2615-2859, 2674-2893, 2822-2989

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
30/1629617CB1/ 3150	1-182, 4-2567, 103-365, 103-825, 104-450, 227-707, 253-891, 743-1064, 819-1193, 853-1162, 853-1216, 987-1401, 987-1401, 987-1408, 987-1418, 987-1422, 988-1197, 988-1399, 988-1400, 988-1401, 988-1414, 988-1416, 989-1114, 989-1401, 1006-1401, 1054-1401, 1063-1516, 1117-1700, 1126-1419, 1126-1563, 1266-1684, 1314-1866, 1410-1637, 1475-1983, 1509-1969, 1541-1985, 1582-1987, 1857-2127, 1916-2199, 2026-2187, 2037-2283, 2046-2568, 2084-2372, 2188-2461, 2188-2742, 2319-2580, 2333-2564, 2359-2572, 2583-2738, 2583-3150
31/2369279CB1/ 2012	1-405, 348-772, 351-405, 564-786, 564-1069, 804-1287, 804-1470, 817-1166, 817-1297, 873-1629, 1101-1443, 1231-1896, 1231-1907, 1231-1960, 1231-1994, 1231-2011, 1231-2012
32/3167506CB1/ 1606	1-1601, 147-443, 148-725, 229-605, 240-606, 297-1606, 356-605, 393-745, 571-796, 575-796
33/3075937CB1/ 7364	1-563, 1-566, 1-615, 1-621, 1-678, 1-818, 1-7347, 2-307, 58-761, 62-601, 96-806, 376-1254, 383-698, 383-704, 412-1254, 433-1254, 455-1180, 462-1254, 469-1254, 549-7364, 851-1263, 1705-1950, 1905-2233, 2023-2633, 2202-2700, 2208-3023, 2220-2682, 2411-2700, 2442-3074, 2631-3281, 2632-3139, 2635-2884, 2639-2924, 2778-3211, 2813-3370, 2891-3501, 2897-3348, 2918-3718, 3038-3485, 3071-3532, 3075-3416, 3331-3834, 3491-4103, 3666-4046, 3666-4047, 3687-4046, 3687-4047, 3691-4005, 3691-4046, 3691-4047, 3815-4046, 3823-4046, 4219-4470, 4219-4721, 4221-4887, 4222-4906, 4224-4897, 4235-4500, 4269-4943, 4308-4964, 4335-4976, 4383-4506, 4392-5038, 4491-5118, 4506-4651, 4539-4781, 4560-5078, 4594-4791, 4652-4781, 4652-4921, 4674-5248, 4676-5245, 4699-5340, 4700-5517, 4755-5324, 4781-5416, 4782-4921, 4782-5010, 4782-5045, 4810-5358, 4828-5357, 4842-5382, 4875-5403, 4890-5327, 4922-5045, 4922-5204, 4955-5576, 4958-5471, 4969-5630, 4977-5267, 4988-5612, 4997-5573, 5002-5582, 5011-5600, 5037-5702, 5044-5403, 5045-5204, 5046-5227, 5049-5612, 5070-5360, 5070-5501, 5082-5345, 5111-5576, 5149-5369, 5165-5732, 5180-5636, 5187-5748, 5188-5875, 5196-5756, 5199-5558, 5217-5779, 5227-5396, 5231-5367, 5231-5459, 5237-5630, 5247-5732, 5249-5538, 5249-5912, 5291-5742, 5332-5598, 5369-5692, 5378-5847, 5381-6053, 5465-5735, 5468-6061, 5494-6093, 5519-5973, 5520-5757, 5534-6071, 5559-5786, 5568-6096, 5573-6094, 5613-6096, 5647-5886, 5660-5920, 5665-6094, 5681-6096, 5702-5952, 5730-6058, 5737-5980, 5743-6045, 5767-6019, 5767-6027, 5783-6045, 5785-6032, 5785-6064, 5785-6282, 5795-6052,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
33	5805-6085, 5810-6042, 5814-6576, 5849-6116, 5902-6176, 5906-6165, 5934-6181, 5952-6225, 6187-6663, 6345-6852, 6377-6660, 6379-6638, 6410-6918, 6416-6627, 6416-6821, 6419-6691, 6419-6722, 6440-6724, 6452-7281, 6458-7026, 6464-6728, 6470-6920, 6470-6981, 6474-6739, 6482-6919, 6484-6735, 6487-6729, 6499-6751, 6499-6759, 6502-7062, 6524-6844, 6535-6952, 6536-6791, 6543-6800, 6549-6757, 6564-6826, 6588-6816, 6602-7278, 6629-7193, 6652-6934, 6661-7284, 6668-7060, 6672-7325, 6679-7014, 6743-7311, 6751-7013, 6751-7022, 6760-7018, 6760-7053, 6764-7333, 6775-7331, 6836-6858, 6847-7306, 6849-7364, 6868-7335, 6885-7140, 6896-7355, 6897-7347, 6927-7364, 6928-7214, 6932-7188, 6937-7204, 6974-7212, 6978-7254, 7009-7287, 7026-7364, 7052-7342, 7067-7317, 7068-7328, 7068-7332, 7068-7341, 7082-7358, 7086-7364, 7100-7361
34/5176268CB1/ 1345	1-281, 1-599, 1-1162, 249-792, 298-580, 298-891, 417-656, 420-841, 542-1151, 876-1345
35/72830854CB1/ 4255	1-370, 1-453, 1-466, 1-482, 1-485, 1-510, 1-516, 1-523, 1-535, 1-560, 1-568, 1-582, 5-453, 5-598, 51-628, 54-574, 114-609, 146-478, 254-954, 278-864, 287-946, 315-953, 351-587, 351-616, 351-628, 351-621, 359-598, 359-1046, 372-1036, 375-997, 377-648, 377-650, 427-1003, 449-572, 466-927, 486-1032, 493-598, 493-607, 493-616, 493-702, 493-992, 537-1139, 538-1139, 574-620, 574-658, 574-692, 575-656, 583-620, 596-692, 603-653, 610-662, 610-721, 610-725, 610-728, 612-728, 621-728, 622-723, 632-1237, 647-697, 647-717, 647-722, 647-727, 647-728, 648-728, 657-728, 657-1393, 668-728, 670-1481, 682-728, 684-728, 687-1254, 693-728, 695-1373, 696-1370, 702-1440, 702-1464, 712-1390, 733-1410, 746-866, 755-1561, 763-1490, 765-1363, 775-1377, 780-1413, 788-1473, 795-1281, 796-1431, 796-1636, 835-1417, 848-1449, 855-1561, 901-1458, 906-1533, 921-1368, 930-1420, 950-1492, 952-1758, 969-1617, 969-1642, 973-1652, 975-1625, 977-1655, 978-1620, 985-1626, 985-1664, 991-1749, 992-1654, 1010-1696, 1021-1559, 1025-1480, 1039-1752, 1042-1582, 1043-1640, 1049-1640, 1050-1643, 1055-1473, 1066-1784, 1068-1737, 1069-1755, 1071-1738, 1085-1679, 1099-1663, 1106-1665, 1108-1676, 1114-1760, 1132-1661, 1156-1681, 1158-1755, 1173-1760, 1179-1623, 1184-1739, 1190-1723, 1206-1712, 1213-1784, 1254-1755, 1269-1758, 1299-1840, 1301-1717, 1305-1707, 1307-1717, 1410-1840,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
35	1413-1984, 1422-1746, 1446-1840, 1507-2176, 1522-2176, 1541-1607, 1557-2166, 1697-2259, 1820-2282, 2031-2440, 2044-2219, 2044-2274, 2044-2319, 2044-2400, 2044-2426, 2044-2442, 2044-2449, 2044-2450, 2044-2519, 2044-2520, 2044-2529, 2044-2547, 2044-2553, 2044-2562, 2044-2613, 2044-2666, 2044-2694, 2044-2704, 2044-2714, 2044-2716, 2044-2721, 2044-2724, 2044-2745, 2044-2756, 2044-2769, 2044-2788, 2044-2811, 2044-2840, 2044-2881, 2047-2824, 2048-2785, 2070-2802, 2078-2695, 2102-2290, 2102-2318, 2102-2319, 2131-2855, 2216-3668, 2223-2733, 2274-2496, 2279-2475, 2314-3110, 2381-2575, 2391-2626, 2486-3135, 2612-3256, 2623-2890, 2662-3338, 2678-3483, 2691-3136, 2702-3403, 2709-3396, 2754-3413, 2756-3372, 2764-3483, 2870-3369, 2884-3460, 2892-3500, 2966-3630, 2979-3171, 2984-3255, 2984-3492, 2995-3237, 2997-3453, 2997-3588, 3019-3236,
	3027-3212, 3118-3625, 3132-3668, 3158-3668, 3164-3415, 3247-3668, 3298-3668, 3375-3916, 3380-3663, 3382-3648, 3385-3682, 3390-3761, 3399-3959, 3403-3720, 3424-3890, 3424-3901, 3432-3856, 3447-3746, 3452-3787, 3490-3668, 3511-3766, 3530-3906, 3616-3725, 3620-3989, 3625-4065, 3628-3857, 3665-3853, 3665-3861, 3665-3903, 3665-3985, 3668-4101, 3668-4233, 3668-4251, 3669-3715, 3669-3754, 3669-3778, 3669-3855, 3669-3874, 3669-3875, 3669-4003, 3669-4120, 3669-4176, 3669-4193, 3669-4226, 3669-4246, 3669-4247, 3669-4251, 3669-4254, 3669-4255, 3670-3903, 3670-3914, 3670-4003, 3670-4036, 3670-4221, 3670-4250, 3687-4253, 3689-4234, 3690-4254, 3708-4251, 3738-4253, 3749-4017, 3756-4024, 3763-3907, 3790-4235, 3829-4234, 3851-4234, 3857-4234, 3892-4234, 3937-4099, 3946-4232, 3977-4153
36/3632052CB1/ 4181	1-352, 353-706, 408-933, 415-926, 426-696, 506-783, 507-703, 507-889, 507-894, 507-950, 507-965, 508-920, 508-960, 508-964, 509-830, 514-692, 514-696, 514-752, 514-841, 514-857, 514-860, 514-923, 514-974, 530-1046, 536-861, 539-638, 556-1310, 565-764, 581-1190, 587-1129, 587-1247, 590-1020, 590-1162, 620-1255, 657-1203, 704-1299, 730-1196, 734-1175, 782-1403, 799-1404, 840-1095, 848-1358, 849-1403, 854-1132, 871-1119, 901-1576, 946-1367, 1117-1358, 1133-1577, 1144-1447, 1144-1477, 1160-1577, 1294-1577, 1320-1570, 1323-1562, 1329-1577, 1342-1577, 1379-1577, 1385-1577, 1386-1577, 1399-1577, 1404-2546, 1412-1577, 1422-1577, 1431-1577, 1775-1823, 1775-1933, 1775-2186, 2219-2635, 2248-2635, 2316-2635, 2355-2635, 2432-2610, 2468-2849, 2525-3823, 2682-2835, 2682-3121, 2682-3269, 2682-3277, 2682-3310, 2682-3325, 2682-3326, 2682-3340, 2682-3444, 2682-3448, 2686-3488, 2818-3631, 2837-3553, 2914-3720, 2941-3687, 2992-3382, 3055-3441, 3109-3940, 3185-3860, 3207-4181

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
37/7493817CB1/ 2478	1-819, 1-2478, 290-464, 290-491, 290-776, 290-829, 290-853, 290-891, 290-901, 290-910, 290-929, 290-931, 290-961, 290-981, 293-914, 327-872, 413-891, 424-927, 450-873, 489-751, 610-1054, 614-1069, 840-1085, 840-1141, 891-1614, 909-1612, 959-1612, 1015-1216, 1015-1545, 1367-1467, 1674-1973
38/6715627CB1/ 2928	1-481, 11-669, 203-809, 226-743, 242-1102, 254-959, 278-1045, 358-994, 359-383, 359-1041, 491-639, 491-674, 491-937, 523-777, 558-936, 558-1194, 558-1274, 559-930, 666-1160, 702-1089, 727-1492, 759-1386, 768-1026, 768-1236, 838-1397, 872-1124, 880-1459, 900-1146, 937-1143, 937-1358, 955-1494, 975-1674, 992-1291, 995-1594, 1136-1293, 1168-1829, 1200-1890, 1221-1889, 1224-1980, 1227-1905, 1234-1416, 1293-1567, 1294-1541, 1297-1436, 1298-1591, 1337-1777, 1358-1635, 1359-1489, 1360-1427, 1370-1589, 1370-1870, 1395-1668, 1395-2171, 1409-1662, 1410-2058, 1433-1875, 1438-2172, 1447-2184, 1454-2081, 1455-1731, 1470-1690, 1472-2219, 1473-2185, 1486-2144, 1516-2170, 1530-2150, 1535-2059, 1603-2321, 1643-2205, 1648-2085, 1650-1839, 1653-2370, 1666-2307, 1725-2354, 1806-2354, 1809-2379, 1809-2388, 1814-2308, 1886-2019, 1891-2388, 1965-2057, 2049-2270, 2115-2388, 2138-2343, 2282-2786, 2310-2928, 2359-2902, 2370-2598, 2454-2485, 2568-2809, 2568-2902
39/7727886CB1/ 3486	1-258, 1-596, 246-974, 327-580, 580-731, 580-1787, 707-956, 713-1327, 770-1365, 866-1478, 994-1620, 1018-1326, 1102-1756, 1265-1760, 1308-1882, 1336-1616, 1394-1970, 1501-2136, 1510-2033, 1516-2145, 1585-2117, 1630-2244, 1825-2237, 1827-2274, 1831-1981, 1835-2206, 1860-2285, 1912-2161, 2040-2310, 2049-2293, 2076-2623, 2141-2414, 2142-2607, 2240-2511, 2244-2547, 2286-2531, 2317-2451, 2317-2573, 2317-2849, 2320-2663, 2321-2792, 2352-2846, 2352-2956, 2352-3013, 2354-2592, 2401-2600, 2401-2721, 2412-2992, 2448-2709, 2469-2586, 2542-3110, 2554-2793, 2573-2817, 2618-3132, 2683-3132, 2683-3292, 2683-3298, 2711-3003, 2715-3315, 2749-3276, 2756-3007, 2760-3030, 2774-3439,
	2786-3086, 2787-3485, 2808-3486, 2824-3086, 2829-3486, 2856-3092, 2856-3110, 2857-3090, 2857-3217, 2858-3120, 2862-3204, 2882-3486, 2903-3484, 2920-3486, 2927-3484, 2944-3438, 2951-3379, 2968-3486, 2976-3189, 2978-3247, 2978-3486, 2982-3109, 2986-3400, 2986-3486, 2987-3190, 2988-3486, 3028-3486, 3029-3486, 3045-3486, 3054-3486, 3059-3486, 3065-3354, 3069-3486, 3075-3340, 3076-3486, 3078-3486, 3098-3486, 3100-3315, 3102-3486, 3104-3486, 3105-3486, 3106-3486, 3108-3486, 3113-3365, 3123-3486, 3140-3438, 3159-3486, 3160-3414, 3177-3486, 3182-3486, 3183-3433, 3183-3439, 3191-3486, 3212-3486, 3230-3280, 3232-3360, 3236-3486, 3259-3486, 3281-3486, 3283-3486, 3289-3468, 3311-3484, 3311-3486, 3312-3486, 3394-3484, 3394-3486

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
40/914113CB1/ 2535	1-623, 5-570, 222-1042, 299-878, 306-865, 319-1000, 362-713, 370-1067, 389-825, 389-1054, 410-738, 410-1097, 510-630, 510-1049, 511-758, 511-789, 511-823, 511-848, 511-864, 511-871, 511-875, 511-885, 511-937, 511-976, 511-981, 511-993, 511-1045, 511-1079, 511-1142, 511-1154, 511-1165, 511-1168, 511-1186, 511-1220, 511-1241, 511-1242, 511-1286, 511-1291, 511-1319, 511-1325, 511-1366, 511-1387, 511-1434, 511-1445, 511-1453, 511-1491, 513-1093, 516-868, 522-1079, 525-602, 534-1441, 594-1286, 626-1505, 662-1505, 721-1464, 776-1505, 788-1366, 817-1430, 833-1435, 836-1469, 892-1503, 896-1416, 911-1387, 911-1505, 913-1273, 918-1505, 920-1496, 921-1503, 923-1491, 923-1500,
	924-1502, 924-1503, 924-1511, 928-1505, 929-1503, 929-1505, 934-1503, 935-1503, 937-1503, 945-1505, 960-1505, 967-1503, 969-1505, 970-1500, 974-1503, 975-1503, 995-1505, 1000-1500, 1001-1503, 1001-1504, 1011-1472, 1011-1511, 1029-1507, 1036-1503, 1052-1500, 1059-1505, 1065-1503, 1073-1503, 1075-1505, 1078-1488, 1103-1503, 1105-1303, 1107-1500, 1108-1503, 1108-1511, 1115-1503, 1120-1503, 1134-1472, 1142-1482, 1156-1503, 1161-1503, 1178-1511, 1181-1788, 1183-1500, 1189-1506, 1193-1435, 1195-1511, 1215-1503, 1236-1486, 1243-1856, 1254-1480, 1274-2094, 1334-1511, 1402-1816, 1403-1816, 1513-1785, 1671-1859, 1714-1962, 1805-1962, 1812-2467, 1830-2027, 2014-2230, 2014-2231, 2014-2235, 2014-2280, 2014-2297, 2014-2300, 2014-2309, 2014-2345, 2014-2359,
	2014-2385, 2014-2427, 2014-2477, 2014-2486, 2014-2487, 2014-2489, 2014-2490, 2014-2491, 2014-2492, 2014-2495, 2014-2496, 2014-2498, 2016-2100, 2016-2268, 2016-2394, 2016-2489, 2017-2508, 2017-2517, 2018-2507, 2020-2525, 2021-2273, 2021-2342, 2021-2524, 2021-2533, 2022-2509, 2022-2520, 2022-2523, 2022-2524, 2026-2441, 2026-2506, 2032-2285, 2032-2373, 2032-2389, 2032-2463, 2034-2253, 2036-2267, 2036-2524, 2041-2463, 2042-2494, 2044-2338, 2044-2495, 2049-2503, 2054-2231, 2055-2226, 2056-2486, 2062-2455, 2084-2433, 2091-2535, 2103-2308, 2104-2489, 2118-2308, 2123-2427, 2139-2307, 2166-2516, 2172-2515, 2186-2523, 2210-2489, 2231-2533, 2290-2524, 2307-2370, 2344-2511

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
41/1953711CB1/ 5514	1-403, 33-491, 261-842, 261-853, 261-878, 290-668, 594-867, 751-1247, 810-1510, 824-1507, 884-1510, 1103-1329, 1277-1845, 1381-1406, 1428-1648, 1428-1920, 1428-1954, 1492-2036, 1498-2036, 1509-2036, 1518-2036, 1540-2036, 1563-2036, 1574-2036, 1575-2028, 1580-2036, 1581-1890, 1581-1894, 1614-1890, 1616-2036, 1632-1787, 1642-2036, 1682-2036, 1693-2036, 1791-2338, 1834-2233, 1840-2362, 1854-2396, 1923-2153, 2002-2569, 2005-2174, 2040-2552, 2041-2635, 2057-2512, 2147-2419, 2163-2688, 2189-2374, 2240-2751, 2261-2759, 2404-2515, 2407-2649, 2420-3042, 2502-2761, 2509-2949, 2658-3191, 2660-3273, 2660-3287, 2698-3034, 2698-3298, 2698-3307, 2743-3333, 2805-3215;
	2837-3217, 2865-3411, 2959-3468, 2979-3468, 2980-3314, 3023-3468, 3028-3461, 3030-3468, 3075-3468, 3178-3217, 3178-3468, 3180-3333, 3199-3468, 3294-3686, 3303-3452, 3373-3810, 3428-3934, 3521-4071, 3537-4140, 3660-3925, 3672-4346, 3801-4030, 3801-4087, 3801-4356, 3802-4356, 3807-4471, 3813-4394, 3865-4192, 3886-4471, 4101-4605, 4101-4620, 4139-4400, 4139-4653, 4361-4479, 4361-4600, 4361-4718, 4361-4744, 4361-4752, 4361-4773, 4361-4774, 4361-4790, 4361-4812, 4361-4834, 4361-4851, 4361-4858, 4361-4898, 4361-4974, 4361-5033, 4362-4633, 4363-4501, 4369-4734, 4369-5056, 4388-4772, 4404-5042, 4603-4788, 4619-5187, 4623-4870, 4623-4903, 4673-5311, 4693-5305, 4794-5027, 4794-5052, 4794-5055, 4818-4978, 4819-5313, 4856-5272, 5221-5359, 5221-5514, 5225-5514
42/1595275CB1/ 3715	1-634, 1-2199, 1-2387, 3-716, 21-768, 35-555, 35-632, 62-460, 126-660, 200-718, 201-898, 206-875, 240-788, 244-941, 250-913, 328-1014, 333-1014, 341-1077, 377-950, 437-582, 457-934, 464-881, 465-917, 487-1138, 519-939, 521-934, 577-626, 585-1077, 617-836, 617-1154, 617-1183, 633-1016, 643-1112, 651-1119, 667-1011, 670-962, 766-1160, 893-1335, 894-1344, 952-1170, 1032-1259, 1215-1323, 1232-1754, 1344-1929, 1374-1746, 1490-2195, 1638-1806, 1713-1862, 1786-2389, 1822-2389, 1852-2516, 1957-2382, 2014-2751, 2084-2393, 2106-2792, 2108-2511, 2182-2399, 2372-2928, 2372-3084, 2507-3109, 2639-3317, 2649-3160, 2649-3342, 2697-3334, 2852-3575, 2919-3715
43/72332548CB1/ 1082	1-230, 1-301, 1-389, 1-414, 1-428, 1-473, 1-474, 1-479, 1-523, 1-680, 2-249, 2-519, 2-523, 3-521, 3-523, 6-369, 6-431, 6-520, 6-523, 12-393, 15-262, 15-268, 15-281, 15-686, 15-692, 15-723, 15-726, 15-729, 16-262, 16-273, 16-330, 120-342, 248-500, 269-447, 299-474, 496-1082, 526-1082, 547-692, 613-1073

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
44/7322834CB1/ 2621	1-367, 24-854, 59-631, 87-569, 115-898, 127-689, 150-792, 260-1041, 275-986, 279-1079, 340-993, 340-1015, 340-1031, 359-912, 359-1018, 365-990, 365-995, 369-1131, 394-1125, 396-916, 441-1103, 479-1255, 487-1015, 513-903, 550-1255, 563-1233, 564-1226, 569-1213, 580-1253, 582-1006, 592-1259, 602-1329, 603-961, 617-1068, 696-971, 708-1381, 754-1035, 758-1018, 771-991, 776-1390, 791-1576, 804-1025, 806-1082, 809-1131, 822-1423, 828-1490, 836-1079, 845-1425, 845-1608, 846-956, 854-1142, 855-1606, 860-1158, 866-1484, 876-1611, 879-1118, 882-1110, 890-1544, 908-1574, 910-1606, 912-1239, 926-1595, 930-1517, 931-1598, 935-1226, 953-1245, 955-1598, 961-1634, 974-1398, 980-1277, 1008-1576, 1030-1486, 1039-1622, 1053-1457, 1087-1357, 1089-1371,
	1091-1698, 1100-1272, 1106-1413, 1106-1585, 1132-1713, 1146-1423, 1149-1332, 1159-1422, 1162-1499, 1162-1566, 1168-1411, 1168-1417, 1168-1423, 1177-1750, 1187-1634, 1191-1728, 1227-1619, 1242-1536, 1249-1486, 1250-1435, 1271-1542, 1366-1764, 1482-2035, 1494-2126, 1642-2134, 1647-2134, 1663-2134, 1682-2134, 1687-2108, 1690-2084, 1694-1965, 1695-2084, 1696-2084, 1704-2134, 1710-2084, 1726-2116, 1733-2134, 1737-2086, 1763-2084, 1777-1961, 1783-2106, 1784-2301, 1793-2420, 1802-2602, 1804-2130, 1811-1930, 1834-2609, 1843-2116, 1892-2069, 1896-2612, 1901-2612, 1904-2621, 1906-2084, 1917-2126, 1918-2621, 1960-2621, 2144-2621, 2197-2522, 2245-2489, 2272-2460, 2350-2514
45/1925714CB1/ 5539	1-419, 101-5539, 201-4986, 2060-2517, 2060-2633, 2060-5071, 2291-2707, 2291-2815, 2291-2887, 2291-3067, 2291-3083, 2292-2952, 2293-3103, 2294-3175, 2294-3283, 2296-2399, 2296-2435, 2296-3226, 2297-2734, 2297-2751, 2297-2827, 2297-2923, 2297-2935, 2297-2936, 2297-3211, 2305-2470, 2332-2471, 2332-3175, 2332-3262, 2366-2700, 2366-3355, 2402-2481, 2402-3283, 2441-2615, 2441-3355, 2507-2676, 2507-3299, 2512-3355, 2513-2578, 2545-2827, 2545-3355, 2561-2611, 2579-2748, 2579-2916, 2579-3355, 2621-2784, 2621-3260, 2652-2820, 2652-2831, 2652-3312, 2693-2856, 2693-2867, 2693-3331, 2708-2734, 2729-2896, 2729-3355, 2759-2928, 2759-3355, 2801-2963, 2801-3331, 2831-3034, 2831-3355, 2873-3046, 2873-3327, 2909-3346, 2939-3355, 2989-3144, 2989-3154,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
45	3197-3262, 3269-3522, 3349-3800, 3349-3991, 3612-4036, 3669-4105, 3798-4207, 3825-4325, 3885-4400, 3901-4150, 3922-4180, 3933-4197, 3962-4461, 3983-4592, 4026-4392, 4131-4645, 4167-4672, 4172-4510, 4230-4672, 4246-4936, 4247-4672, 4249-4672, 4265-4672, 4342-4602, 4369-5019, 4375-4916, 4378-4865, 4453-4746, 4453-4853, 4456-4914, 4464-5006, 4465-4716, 4510-4672, 4517-4768, 4540-5218, 4600-4824, 4709-5004, 4709-5012, 4725-5137, 4802-5075, 4843-5102, 4844-5138, 4844-5313, 4881-5497, 4891-5140, 5068-5523, 5071-5494, 5085-5348, 5087-5319, 5087-5336, 5092-5349, 5092-5380, 5093-5531, 5097-5361, 5099-5300, 5099-5355, 5108-5526, 5110-5318, 5138-5366, 5148-5352, 5160-5537, 5203-5455, 5236-5537, 5237-5535, 5239-5537, 5254-5525, 5264-5537, 5285-5523, 5322-5537, 5369-5515, 5369-5523, 5383-5535
46/6803363CB1/ 2338	1-707, 39-2338, 51-612, 190-845, 282-589, 599-1373, 608-1234, 914-1504, 957-1504, 1036-1500, 1276-1513, 1297-1474, 1757-2089, 1763-2327
47/7070580CB1/ 3214	1-3190, 176-680, 448-855, 497-701, 519-815, 572-1164, 585-842, 585-984, 585-1005, 585-1043, 680-1171, 767-1401, 815-1377, 877-1108, 877-1469, 877-1471, 877-1493, 894-1129, 1045-1186, 1047-1173, 1098-1654, 1098-1748, 1144-1485, 1188-1815, 1198-1752, 1218-1485, 1262-1901, 1273-1598, 1316-1914, 1324-1936, 1325-1960, 1348-2030, 1364-1880, 1371-1685, 1371-1703, 1373-1947, 1385-1944, 1407-2057, 1418-2027, 1432-1911, 1459-1805, 1469-2027, 1479-1864, 1528-1881, 1558-2179, 1562-2126, 1570-2195, 1607-2274, 1613-2208, 1617-2147, 1620-2263, 1627-2195, 1641-2208, 1642-1984, 1646-1826, 1646-2245, 1654-2282, 1658-2127, 1669-2337, 1671-2173, 1673-2196, 1677-2264, 1686-2312, 1690-1916, 1690-2082, 1699-2189, 1712-2405, 1729-2270, 1734-2201, 1741-2354, 1749-2265,
	1751-2378, 1754-2034, 1757-2392, 1764-2388, 1773-1980, 1816-2363, 1829-2492, 1839-2386, 1846-2467, 1847-2465, 1858-2239, 1874-2465, 1876-2123, 1892-2535, 1897-2415, 1900-2411, 1908-2526, 1934-2307, 1934-2393, 1934-2406, 1934-2414, 1934-2554, 1939-2545, 1950-2499, 1958-2507, 1958-2533, 1960-2625, 1971-2264, 1971-2554, 1983-2625, 1988-2526, 2004-2668, 2044-2288, 2044-2300, 2044-2572, 2055-2676, 2057-2649, 2060-2658, 2065-2667, 2066-2768, 2081-2631, 2092-2622, 2124-2623, 2124-2666, 2126-2747, 2134-2761, 2140-2782, 2153-2630, 2153-2655, 2193-2741, 2195-2556, 2205-2768, 2211-2620, 2288-2742, 2296-2535, 2298-2535, 2502-3079, 2553-3046, 2587-3062, 2591-3214, 2600-3185, 2625-3077, 2698-3051, 2701-2953, 2701-3162, 2701-3188, 2741-3189, 2744-2981, 2851-3103, 2851-3106, 2851-3168, 2889-3169, 2894-3171

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
48/7500176CB1/ 2824	1-271, 1-2791, 126-359, 154-772, 157-714, 170-474, 170-624, 170-666, 170-749, 170-759, 172-738, 172-793, 174-563, 174-730, 176-428, 177-555, 177-661, 177-715, 177-787, 181-525, 183-576, 184-421, 185-719, 185-788, 186-761, 187-293, 187-342, 187-492, 187-638, 188-639, 189-518, 194-673, 195-315, 195-434, 195-582, 198-525, 220-688, 249-570, 294-526, 335-1046, 350-996, 364-982, 395-1266, 408-867, 464-744, 511-1196, 542-1093, 545-985, 545-1034, 547-1096, 550-945, 561-1286, 590-1303, 594-1278, 599-738, 604-1135, 612-1156, 622-1294, 625-1254, 626-1231, 628-1092, 637-1149, 645-1173, 646-1086,
	681-1177, 681-1324, 685-1263, 690-1106, 695-1421, 696-1252, 698-1311, 719-1482, 727-1228, 734-1283, 734-1334, 772-1330, 774-1267, 777-906, 785-1343, 790-1408, 812-1381, 813-1252, 819-1443, 834-1572, 850-1443, 856-1086, 866-1031, 866-1516, 886-1486, 892-1122, 892-1316, 897-1232, 900-1124, 910-1538, 913-1231, 914-1462, 915-1191, 918-1327, 926-1394, 938-1519, 944-1148, 951-1246, 951-1632, 955-1200, 960-1244, 961-1209, 962-1227, 964-1316, 969-1249, 970-1241, 974-1259, 976-1682, 978-1208, 978-1250, 986-1277, 988-1687, 989-1276, 1014-1595, 1020-1275, 1020-1293, 1030-1280,
	1041-1166, 1050-1317, 1060-1292, 1063-1351, 1074-1409, 1078-1556, 1083-1356, 1091-1379, 1099-1367, 1118-1343, 1118-1385, 1125-1801, 1127-1350, 1134-1317, 1150-1387, 1152-1426, 1155-1421, 1160-1496, 1176-1712, 1178-1379, 1196-1471, 1197-1531, 1215-1922, 1225-1844, 1226-1511, 1226-1530, 1237-1512, 1308-1501, 1312-1561, 1332-1665, 1337-1823, 1344-1823, 1350-1589, 1523-1766, 1523-2047, 1527-1696, 1556-2032, 1559-2047, 1563-1813, 1575-1744, 1575-1762, 1582-2131, 1584-2120, 1586-1809, 1607-1874, 1611-2253, 1622-1926, 1622-1974, 1628-1885, 1628-1903, 1634-1921, 1640-1862, 1652-1922, 1653-2119, 1655-1905, 1655-1928, 1662-1906, 1663-1905, 1663-1925, 1663-1926, 1664-1922, 1664-1924, 1665-1952, 1670-1969, 1673-1796, 1673-1901, 1673-1947, 1678-1931, 1678-1964, 1679-1911, 1681-2159, 1683-1958, 1684-1956, 1687-1888, 1689-1896, 1689-1980,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
48	1701-1969, 1704-1944, 1711-1903, 1716-1861, 1735-2021, 1754-2019, 1757-1998, 1761-1995, 1768-2029, 1797-2050, 1801-2546, 1820-2037, 1827-2632, 1831-2530, 1860-2730, 1872-2130, 1888-2574, 1892-2137, 1929-2724, 1933-2728, 1936-2416, 1939-2741, 1945-2519, 1949-2384, 1949-2751, 1963-2726, 1966-2052, 1966-2076, 1971-2724, 1991-2725, 2003-2757, 2005-2724, 2014-2459, 2024-2496, 2024-2728, 2033-2660, 2034-2572, 2037-2537, 2038-2727, 2042-2729, 2044-2687, 2045-2456, 2046-2725, 2047-2505, 2064-2758, 2066-2639, 2075-2383, 2079-2438, 2079-2745, 2081-2771, 2082-2435, 2084-2597, 2091-2565, 2091-2724, 2092-2726, 2095-2754, 2097-2697, 2099-2726, 2099-2752, 2100-2756, 2108-2667, 2108-2715, 2108-2726, 2112-2717, 2114-2623, 2114-2752, 2115-2718, 2115-2760, 2120-2416, 2121-2765, 2126-2445, 2130-2576, 2133-2426, 2134-2381, 2134-2384, 2135-2372, 2138-2388, 2139-2375, 2139-2458, 2141-2424, 2141-2427, 2142-2391, 2142-2443, 2143-2419,
	2147-2787, 2149-2701, 2152-2686, 2155-2341, 2155-2393, 2155-2428, 2157-2438, 2158-2414, 2160-2421, 2160-2774, 2163-2429, 2163-2622, 2164-2404, 2164-2429, 2165-2428, 2167-2756, 2170-2728, 2172-2443, 2176-2413, 2176-2418, 2177-2327, 2177-2399, 2177-2455, 2178-2678, 2182-2624, 2183-2784, 2184-2422, 2185-2785, 2186-2485, 2188-2451, 2188-2477, 2188-2728, 2191-2481, 2196-2528, 2196-2706, 2199-2490, 2201-2398, 2203-2478, 2203-2580, 2205-2432, 2208-2667, 2208-2747, 2210-2401, 2221-2443, 2223-2521, 2227-2752, 2228-2472, 2230-2479, 2233-2471, 2237-2446, 2240-2746, 2243-2726, 2244-2667, 2246-2743, 2248-2449, 2249-2666, 2249-2779, 2251-2530, 2251-2545, 2254-2505, 2258-2685, 2259-2689, 2261-2470, 2262-2531, 2264-2546, 2264-2690, 2266-2718, 2266-2725, 2269-2534, 2269-2547, 2270-2749, 2275-2515, 2277-2582, 2278-2542, 2280-2547, 2287-2583, 2287-2703, 2298-2559, 2298-2788, 2302-2545, 2304-2493, 2304-2527, 2304-2543, 2307-2490, 2307-2549,
	2307-2824, 2308-2769, 2314-2612, 2315-2509, 2315-2564, 2318-2790, 2325-2587, 2325-2614, 2325-2725, 2325-2743, 2325-2752, 2325-2753, 2326-2696, 2327-2726, 2339-2689, 2346-2743, 2346-2757, 2347-2632, 2347-2642, 2351-2619, 2351-2650, 2360-2522, 2364-2636, 2364-2638, 2368-2619, 2372-2712, 2379-2609, 2379-2744, 2380-2757, 2381-2608, 2381-2627, 2381-2631, 2382-2618, 2392-2645, 2392-2682, 2394-2659, 2400-2598, 2400-2682, 2400-2694, 2403-2656, 2403-2669, 2403-2797, 2410-2757, 2411-2636, 2414-2797, 2418-2635, 2422-2799, 2425-2751, 2434-2797, 2437-2707, 2438-2723, 2440-2692, 2446-2651, 2460-2706, 2461-2719, 2472-2750, 2474-2596, 2482-2706, 2484-2713, 2484-2724, 2497-2760, 2499-2761, 2504-2721, 2510-2797, 2512-2791, 2513-2749, 2518-2757, 2540-2737, 2544-2797, 2545-2757, 2590-2783

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
49/7500506CB1/ 1570	1-225, 46-698, 46-774, 46-982, 48-190, 63-250, 351-917, 352-1063, 389-911, 392-768, 409-1070, 419-968, 432-911, 444-1111, 448-1307, 458-702, 470-854, 493-985, 495-1161, 523-800, 525-993, 527-793, 527-891, 556-658, 573-857, 574-967, 583-1158, 641-945, 653-883, 661-929, 682-979, 684-870, 688-1204, 688-1212, 692-1547, 695-955, 704-1382, 712-975, 747-1015, 747-1269, 788-1547, 803-1199, 838-1313, 860-1132, 865-1076, 872-1143, 880-1182, 883-1250, 919-1547, 945-1248, 985-1236, 991-1562, 1006-1245, 1089-1557, 1097-1570, 1117-1570, 1128-1570, 1135-1570, 1137-1556, 1156-1553, 1298-1557, 1470-1570, 1472-1570
50/7500639CB1/ 725	1-209, 1-605, 4-272, 8-273, 8-278, 8-283, 8-534, 11-273, 18-274, 33-303, 38-328, 43-310, 49-285, 57-454, 115-566, 117-370, 117-382, 120-686, 133-400, 169-725, 172-432, 362-669, 376-690, 566-723
51/7506167CB1/ 3108	1-803, 61-729, 72-3098, 124-853, 124-874, 124-958, 124-1008, 136-576, 297-565, 385-1103, 664-1172, 668-1410, 895-1193, 923-1151, 951-1232, 1009-1118, 1058-1302, 1107-1331, 1129-1410, 1268-1425, 1353-1971, 1356-1880, 1425-1699, 1426-1673, 1469-1909, 1490-1767, 1502-1721, 1502-1971, 1527-1800, 1602-1822, 1657-1971, 1662-1945, 1706-2004, 1764-2116, 1782-2398, 1787-2482, 1799-2004, 1803-2482, 1804-2047, 1862-2251, 1893-2531, 1898-2143, 1968-2400, 1971-2237, 1972-2620, 1979-2508, 1981-2167, 1993-2275, 2011-2525, 2034-2599, 2041-2293, 2052-2273, 2066-2540, 2073-2498, 2074-2524, 2085-2716, 2139-2839, 2141-2346, 2141-2349, 2154-2609, 2158-2354, 2161-2364, 2161-2411, 2167-2682, 2168-2407, 2171-2682, 2184-2400, 2190-2796, 2191-2757, 2198-2678, 2201-2431, 2201-2453, 2211-2490, 2211-2738, 2215-2760, 2217-2448, 2230-2846, 2243-2873, 2245-2861, 2248-2506, 2252-2515, 2252-2558, 2260-2934, 2262-2702, 2279-2503, 2284-2549, 2285-2789, 2286-2851, 2287-2914, 2309-2437, 2313-2933, 2345-2593, 2357-2962, 2396-2854, 2405-2871, 2412-2873, 2414-3063, 2420-2871, 2422-2873, 2426-3072, 2430-2843, 2447-2873, 2449-2860, 2452-2873, 2459-2872, 2461-2870, 2466-2872, 2466-2873, 2467-2873, 2488-3090, 2493-2868, 2493-2871, 2495-2873, 2496-3108, 2501-2871, 2502-2863, 2508-2873, 2548-2864, 2565-2871, 2578-2849, 2581-2801, 2581-2872, 2601-2851, 2614-3095, 2620-2838, 2621-2868, 2623-2871, 2632-2863, 2638-2802, 2641-3024, 2649-3093, 2666-3095, 2668-3007, 2678-3098, 2681-3088, 2691-2944, 2692-3095, 2703-2955, 2731-3005, 2751-2998, 2753-3108, 2762-3091, 2779-2871, 2796-3041, 2807-3099, 2807-3108, 2808-3095, 2813-3089, 2816-3095, 2943-3043
52/9081189CB1/ 1962	1-652, 1-744, 1-1962, 340-1262, 596-1488, 1039-1962, 1071-1962

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
53/7510095CB1/3005	1-271, 1-3005, 126-359, 137-788, 137-848, 137-869, 137-870, 137-880, 142-805, 170-460, 170-474, 170-531, 170-586, 170-666, 170-749, 170-759, 170-772, 170-812, 176-404, 176-428, 176-435, 177-445, 177-555, 177-661, 177-666, 177-776, 180-475, 181-386, 181-406, 181-419, 181-427, 181-441, 181-454, 181-462, 181-468, 181-493, 181-525, 181-721, 182-374, 182-459, 182-475, 182-509, 182-787, 182-866, 183-423, 183-442, 183-576, 183-870, 184-421, 184-438, 184-457, 184-849, 184-870, 185-475, 185-516, 185-643, 185-719, 185-730, 185-746, 185-750, 185-788, 185-805, 185-814, 185-824, 185-849, 185-854, 185-870, 186-438, 186-454, 186-463, 186-621, 186-761, 186-869, 187-266, 187-293, 187-341, 187-371, 187-429, 187-434, 187-435, 187-443, 187-449, 187-450, 187-451, 187-455, 187-460, 187-463, 187-466, 187-475, 187-476, 187-478, 187-479, 187-487, 187-492, 187-511, 187-538, 187-544, 187-556, 187-566, 187-638, 187-643, 187-682, 187-726, 187-728, 187-730, 187-753, 187-787, 187-788, 187-789, 187-824, 187-849, 187-869, 187-870, 187-872, 188-273, 188-306, 188-425, 188-430, 188-432, 188-436, 188-439, 188-447, 188-450, 188-451, 188-456, 188-457, 188-463, 188-485, 188-495, 188-499, 188-553, 188-639, 188-726, 188-793, 188-806, 188-863, 188-870, 189-413, 189-518, 189-661, 189-693, 189-788, 190-417, 190-426, 190-438, 190-447, 190-711, 190-729, 190-774, 190-867, 190-870, 191-469, 191-693, 191-729, 192-419, 192-451, 192-870, 193-385, 193-425, 193-438, 193-450, 193-455, 193-738, 193-749, 193-788, 193-797, 194-321, 194-502, 194-673, 195-315, 195-399, 195-416, 195-427, 195-428, 195-431, 195-434, 195-441, 195-452, 195-453, 195-458, 195-474, 195-478, 195-483, 195-488, 195-498, 195-582, 195-715, 195-718, 195-813, 195-869, 195-870, 195-875, 196-314, 196-434, 196-443, 196-447, 196-452, 196-663, 196-745, 197-446, 197-466, 197-704, 197-740, 197-772, 197-787, 198-435, 198-436, 198-525, 198-630,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
53	198-728, 198-770, 198-782, 199-560, 199-742, 199-861, 200-402, 200-410, 200-468, 200-472, 200-488, 200-686, 200-711, 201-432, 201-490, 201-722, 202-647, 203-446, 203-761, 204-432, 204-491, 204-553, 205-501, 206-872, 207-657, 208-516, 209-824, 211-767, 215-797, 216-655, 218-446, 220-688, 224-511, 226-468, 226-469, 234-516, 234-860, 240-537, 245-798, 249-570, 249-870, 253-507, 255-661, 255-751, 256-842, 260-718, 266-879, 269-794, 269-913, 288-607, 294-526, 297-866, 298-1104, 299-847, 302-539, 308-870, 316-864, 316-869, 322-876, 335-1046, 350-996, 364-982, 370-647, 379-616, 408-867, 410-633, 410-672, 410-978, 410-1003, 438-678, 449-851, 454-654, 455-868, 456-709, 456-720, 457-715, 461-871, 464-742, 464-744, 481-877, 488-1109, 511-1012, 511-1196, 525-1035, 531-789, 538-788, 542-1093, 545-822, 545-887, 545-979, 547-1096, 550-945, 559-909, 577-1169, 577-1328, 589-860, 590-1166, 591-852, 595-874, 599-738, 599-822, 600-886, 604-879, 604-1135, 606-890, 612-1089, 612-1156, 616-930,
	622-1294, 625-864, 625-1166, 626-1231, 627-1172, 628-1092, 634-871, 635-868, 635-915, 635-923, 635-1095, 637-1149, 640-1149, 646-1086, 653-852, 653-907, 658-908, 658-924, 681-1177, 681-1324, 685-1263, 690-1106, 695-1321, 696-1252, 698-1286, 698-1311, 703-937, 707-924, 708-950, 710-904, 722-1140, 723-1168, 727-1228, 734-1283, 745-1156, 748-979, 754-975, 754-994, 754-1021, 754-1304, 759-982, 762-1029, 772-1330, 774-1267, 777-906, 784-1278, 785-1343, 790-1408, 810-940, 812-1381, 813-1252, 819-1443, 850-1443, 856-1086, 866-1031, 885-1516, 886-1486, 892-1122, 892-1316, 897-1232, 898-1163, 900-1124, 900-1741, 910-1170, 910-1432, 913-1231, 914-1462, 915-1191, 918-1327, 921-1334, 926-1394, 938-1195, 938-1519, 941-1568, 944-1173, 951-1246, 955-1200, 960-1244, 961-1209, 962-1227, 964-1316, 969-1249, 970-1241, 979-1538, 979-1557, 980-1455, 981-1265, 982-1211, 983-1208,
	983-1250, 983-1259, 986-1277, 989-1276, 1002-1574, 1004-1586, 1011-1315, 1014-1557, 1020-1275, 1020-1293, 1030-1280, 1036-1610, 1041-1166, 1042-1733, 1050-1317, 1050-1319, 1060-1292, 1060-1317, 1063-1351, 1071-1359, 1074-1409, 1078-1556, 1083-1356, 1091-1379, 1099-1367, 1118-1343, 1118-1385, 1125-1846, 1127-1350, 1127-1351, 1127-1584, 1134-1317, 1150-1387, 1152-1426, 1155-1421, 1163-1533, 1176-1757, 1178-1379, 1178-1450, 1196-1479, 1196-1499, 1209-1488, 1226-1511, 1229-1453, 1252-1561, 1257-1511, 1258-1569, 1263-1481, 1267-1517, 1279-1539, 1279-1602, 1299-1549, 1308-1552, 1312-1541, 1523-1811, 1809-2083, 1825-2135, 1896-2361, 1930-2221, 1930-2511, 2097-2388, 2099-2338, 2099-2355, 2099-2390, 2099-2437, 2100-2350, 2110-2351, 2117-2369, 2128-2357, 2128-2382, 2131-2319, 2131-2355, 2133-2373, 2133-2446, 2136-2571, 2138-2390, 2162-2369, 2179-2461, 2222-2452, 2224-2724, 2437-2666, 2437-2669

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
54/7510096CB1/ 2906	1-271, 1-2906, 186-912, 255-724, 555-1059, 569-1082, 621-1278, 650-934, 679-912, 758-834, 798-1351, 806-1076, 942-1210, 954-1217, 988-1195, 1085-1518, 1301-1927, 1365-1778, 1390-2310, 1391-1933, 1397-1915, 1410-2015, 1414-1941, 1416-1645, 1416-1856, 1425-1661, 1438-1748, 1438-2040, 1466-1922, 1475-1748, 1478-1737, 1478-1769, 1479-1831, 1487-2046, 1495-2105, 1502-1816, 1511-1822, 1521-1793, 1522-1727, 1525-1842, 1528-1781, 1528-1784, 1528-1785, 1528-2107, 1533-2177, 1537-2310, 1542-1842, 1549-1886, 1552-1778, 1553-1785, 1557-1810, 1560-1830, 1567-1901, 1567-1902, 1567-1904, 1567-1996, 1567-1998, 1567-2045, 1567-2053, 1567-2057, 1567-2064, 1567-2078, 1567-2087,
	1567-2089, 1569-2073, 1580-2037, 1581-1819, 1582-2309, 1586-2179, 1590-1829, 1593-1998, 1594-1741, 1613-2087, 1623-1823, 1648-2124, 1655-1905, 1667-1836, 1667-1854, 1674-2223, 1676-2135, 1678-1901, 1694-1795, 1699-1966, 1703-2232, 1714-1842, 1714-2066, 1720-1977, 1720-1995, 1726-2013, 1744-2014, 1745-2211, 1747-1997, 1747-2020, 1754-1998, 1755-1997, 1755-2017, 1755-2018, 1756-2014, 1756-2016, 1757-2044, 1759-2245, 1762-2061, 1765-1888, 1765-1993, 1765-2039, 1770-2023, 1770-2056, 1771-2003, 1771-2369, 1773-2181, 1775-2050, 1776-2048, 1779-1980, 1781-1988, 1781-2072, 1783-2248, 1793-2061, 1796-2036, 1803-1995, 1813-2381,
	1821-2539, 1827-2113, 1835-2139, 1837-2268, 1840-2641, 1841-2132, 1846-2111, 1849-2090, 1853-2087, 1860-2121, 1866-2157, 1866-2161, 1882-2152, 1883-2299, 1889-2099, 1889-2142, 1889-2158, 1893-2638, 1896-2140, 1902-2138, 1902-2167, 1912-2129, 1923-2353, 1923-2470, 1930-2111, 1931-2172, 1931-2191, 1936-2246, 1943-2393, 1951-2196, 1963-2358, 1964-2222, 1967-2232, 1967-2261, 1969-2191, 1974-2417, 1978-2579, 1979-2190, 1980-2666, 1984-2229, 1985-2455, 1992-2245, 1999-2447, 2003-2112, 2004-2243, 2004-2260, 2004-2295, 2004-2342, 2005-2255, 2005-2571, 2012-2257, 2012-2271, 2015-2366, 2018-2304, 2018-2430, 2021-2816, 2028-2508, 2034-2299, 2036-2224, 2036-2260, 2036-2278, 2037-2611, 2038-2307, 2038-2351, 2041-2476, 2043-2279, 2043-2296, 2043-2306, 2043-2354, 2044-2303, 2044-2318, 2049-2273, 2051-2181, 2051-2329, 2055-2833, 2056-2457, 2058-2353, 2061-2310, 2065-2356, 2067-2274, 2071-2568, 2072-2362, 2076-2290, 2083-2387, 2083-2602,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
54	2084-2340, 2084-2346, 2084-2347, 2084-2366, 2086-2336, 2086-2337, 2086-2399, 2089-2335, 2092-2632, 2095-2353, 2095-2849, 2096-2328, 2097-2816, 2102-2368, 2105-2323, 2105-2407, 2106-2551, 2114-2354, 2114-2360, 2116-2588, 2116-2820, 2117-2307, 2122-2382, 2123-2343, 2123-2347, 2125-2752, 2126-2664, 2127-2357, 2129-2629, 2130-2415, 2136-2779, 2137-2403, 2137-2548, 2139-2597, 2140-2425, 2143-2416, 2144-2311, 2144-2369, 2144-2436, 2146-2271, 2146-2392, 2158-2731, 2167-2475, 2171-2418, 2171-2530, 2173-2863, 2174-2527, 2176-2689, 2177-2421, 2183-2657, 2187-2846, 2189-2789, 2191-2844, 2192-2848, 2197-2437, 2197-2462, 2199-2424, 2200-2807, 2204-2809, 2205-2454, 2206-2715, 2207-2824, 2207-2852, 2211-2509, 2212-2508, 2213-2482, 2217-2486, 2217-2819, 2218-2537, 2222-2668, 2225-2518, 2226-2473, 2226-2476, 2227-2462, 2227-2464, 2227-2491, 2227-2531, 2230-2480, 2230-2489, 2231-2467, 2231-2550, 2233-2516, 2233-2519, 2234-2483, 2234-2535, 2235-2511, 2241-2793, 2244-2474, 2244-2479, 2244-2778, 2245-2470, 2247-2433, 2247-2437, 2247-2485, 2247-2520, 2249-2530, 2250-2509, 2252-2485, 2252-2513, 2252-2866, 2255-2481, 2255-2521, 2255-2714, 2256-2476, 2256-2496, 2256-2521, 2257-2520, 2258-2814, 2259-2848, 2262-2821, 2264-2535, 2268-2505, 2268-2510, 2269-2419, 2269-2491, 2269-2547, 2270-2601, 2273-2455, 2274-2716, 2277-2421, 2278-2577, 2280-2543, 2280-2555, 2280-2569, 2280-2820, 2281-2485, 2283-2573, 2288-2620, 2288-2798, 2291-2582, 2293-2490, 2295-2514, 2295-2570, 2295-2601, 2297-2524, 2300-2759, 2302-2493, 2313-2535, 2315-2613, 2319-2844, 2320-2564, 2322-2571, 2323-2851, 2325-2563, 2328-2817, 2329-2538, 2336-2759, 2338-2849, 2340-2541, 2341-2758, 2342-2571, 2342-2574, 2343-2622, 2343-2637, 2346-2597, 2350-2777, 2351-2781, 2353-2562, 2354-2623, 2356-2638, 2356-2782, 2357-2853, 2358-2810, 2361-2626, 2361-2639, 2362-2841, 2365-2648, 2367-2607, 2369-2674, 2370-2634, 2372-2639, 2372-2849, 2379-2675, 2379-2795, 2390-2651, 2394-2637, 2396-2585, 2396-2619, 2396-2635, 2397-2854, 2399-2582, 2399-2641, 2399-2906, 2400-2861, 2406-2704, 2407-2601, 2407-2656, 2410-2871, 2417-2679, 2417-2706, 2417-2817, 2417-2818, 2417-2844, 2417-2845, 2418-2788, 2419-2858, 2438-2849, 2439-2724, 2439-2734, 2443-2711, 2443-2742, 2450-2853, 2452-2614, 2456-2728, 2456-2730, 2460-2711, 2471-2701, 2471-2836, 2472-2849, 2473-2700, 2473-2719, 2473-2723, 2477-2660, 2477-2710, 2484-2774, 2485-2840, 2486-2751, 2489-2748, 2492-2690, 2492-2774, 2492-2786, 2495-2748, 2495-2761, 2502-2849, 2503-2728, 2510-2727, 2517-2843, 2529-2799, 2530-2815, 2532-2784, 2538-2743, 2540-2764, 2550-2781, 2552-2798, 2553-2811, 2564-2842, 2566-2688, 2574-2798, 2576-2805, 2576-2816, 2589-2852, 2591-2853, 2596-2813, 2604-2878, 2605-2841, 2610-2849, 2682-2844

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
28	261510CB1	THYRNOT03
29	7498674CB1	BRAUNOR01
30	1629617CB1	BRSTNOT07
31	2369279CB1	BRAXTDR15
32	3167506CB1	OVARNOE02
33	3075937CB1	BONEUNR01
34	5176268CB1	EPIMNON05
35	728308S4CB1	BRAIFEF02
36	3632052CB1	LIVRNON08
37	7493817CB1	OVARDIJ01
38	6715627CB1	BRAINOY02
39	7727886CB1	BLADTUT02
40	914113CB1	HEAONOR01
41	1953711CB1	BRABDIR01
42	1595275CB1	SINTTMR02
43	7233254SCB1	SCORNON02
44	7322834CB1	KIDETXF04
45	19257714CB1	BRAVUNT02
46	6803363CB1	BRAUNOR01
47	7070580CB1	BRAITUT08
48	7500176CB1	UTRSDIC01
49	7500506CB1	OVARTUE01
50	7500639CB1	BRONNOT02
51	7506167CB1	ISLTNOT01
53	7510095CB1	BRAIFEC01
54	7510096CB1	BRAYDIN03

Table 6

Library	Vector	Library Description
BLADTUT02	pINCY	Library was constructed using RNA isolated from bladder tumor tissue removed from an 80-year-old Caucasian female during a radical cystectomy and lymph node excision. Pathology indicated grade 3 invasive transitional cell carcinoma. Family history included acute renal failure, osteoarthritis, and atherosclerosis.
BONEUNR01	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from an untreated MG-63 cell line derived from an osteosarcoma tumor removed from a 14-year-old Caucasian male (donor A) and using mRNA isolated from sacral bone tumor tissue removed from an 18-year-old Caucasian female (donor B) during an exploratory laparotomy and soft tissue excision. Pathology indicated giant cell tumor of the sacrum in donor B. Donor B's history included pelvic joint pain, constipation, urinary incontinence, unspecified abdominal/pelvic symptoms, and a pelvic soft tissue malignant neoplasm. Family history included prostate cancer in donor B.
BRABDIR01	pINCY	Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease, emphysema, and tobacco abuse.
BRAIFEC01	pINCY	This large size-fractionated library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
BRAIFEF02	PCMV-ICIS	This full-length enriched library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.
BRAINOY02	pINCY	This large size-fractionated and normalized library was constructed using pooled cDNA generated using mRNA isolated from midbrain, inferior temporal cortex, medulla, and posterior parietal cortex tissues removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. Scattered throughout the cerebral cortex, there were multiple small microscopic areas of cavitation with surrounding gliosis. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver. 0.28 million independent clones from this size-selected library were normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAITUT08	pINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 47-year-old Caucasian male during excision of cerebral meningeal tissue. Pathology indicated grade 4 fibrillary astrocytoma with focal tumoral radionecrosis. Patient history included cerebrovascular disease, deficiency anemia, hyperlipidemia, epilepsy, and tobacco use. Family history included cerebrovascular disease and a malignant prostate neoplasm.

Table 6

Library	Vector	Library Description
BRAUNOR01	pINCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles.
		The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloidal goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.
BRAVUNT02	PSPORT1	Library was constructed using pooled RNA isolated from separate populations of unstimulated astrocytes.
BRAXTDR15	PCDNA2.1	This random primed library was constructed using RNA isolated from superior parietal neocortex tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRAYDIN03	pINCY	This normalized library was constructed from 6.7 million independent clones from a brain tissue library. Starting RNA was made from RNA isolated from diseased hypothalamus tissue removed from a 57-year-old Caucasian male who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 -hours/round) reannealing hybridization was used. The library was linearized and recircularized to select for insert containing clones.

Table 6

Library	Vector	Library Description
BRONNOT02	pINCY	Library was constructed using RNA isolated from right lower lobe bronchial tissue removed from a pool of 9 nonasthmatic Caucasian male and female donors, 18- to 55-years-old during bronchial pinch biopsies. Patient history included atopy as determined by positive skin tests to common aero-allergens with no bronchial hyperresponsiveness to histamine. The donors were not current smokers and had no history of alcohol or drug abuse.
BRSTNOT07	pINCY	Library was constructed using RNA isolated from diseased breast tissue removed from a 43-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated mildly proliferative fibrocystic changes with epithelial hyperplasia, papillomatosis, and duct ectasia. Pathology for the associated tumor tissue indicated invasive grade 4, nuclear grade 3 mammary adenocarcinoma with extensive comedo necrosis. Family history included epilepsy, cardiovascular disease, and type II diabetes.
EPIMNON05	pINCY	This normalized mammary epithelial cell tissue library was constructed from 3.28 million independent clones from an epithelial cell tissue library. Starting RNA was made from untreated mammary epithelial cell tissue removed from a 21-year-old female. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 -hours/round) reannealing hybridization was used.
HEAONOR01	pINCY	This random primed library was constructed using pooled RNA isolated from aorta tissue removed from a 10-year-old Caucasian male (donor A) who died from anoxia and a 27-year-old Caucasian female (donor B) who died from an intracranial bleed. Patient history included asthma and suicidal tendency in donor A, and cerebral agenesis in donor B.
ISLTNOT01	pINCY	Library was constructed using RNA isolated from a pooled collection of pancreatic islet cells.
KIDETXF04	PCMV-ICIS	Library was constructed using RNA isolated from a treated, transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue. The cells were treated with 5-aza-2'-deoxycytidine (5AZA) for 72 hours and Trichostatin A for 24 hours and transformed with adenovirus 5 DNA.
LIVRNON08	pINCY	This normalized library was constructed from 5.7 million independent clones from a pooled liver tissue library. Starting RNA was made from pooled liver tissue removed from a 4-year-old Hispanic male who died from anoxia and a 16 week female fetus who died after 16-weeks gestation from anencephaly. Serologies were positive for cytolomegalovirus in the 4-year-old. Patient history included asthma in the 4-year-old. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother of the fetus. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.

Table 6

Library	Vector	Library Description
OVARDIT01	pIGEN	This random primed 5' cap isolated library was constructed using RNA isolated from diseased right ovary tissue removed from a 47-year-old Caucasian female during total abdominal hysterectomy, dilation and curettage, bilateral salpingo-oophorectomy, repair of ureter, and incidental appendectomy. Pathology indicated endometriosis. Pathology for the associated tumor tissue indicated multiple leiomyomata. The left ovary contained a corpus luteum. There was endometriosis involving the posterior serosa. The patient presented with metrorrhagia and a benign neoplasm of the ovary. Patient history included normal delivery, joint pain in multiple joints, and unilateral congenital hip dislocation. Previous surgeries included total hip replacement. Patient medications included calcium. Family history included kidney cancer in the mother; atherosclerotic coronary artery disease and aortocoronary bypass of 3 coronary arteries in the father; benign hypertension and Hodgkin's disease in the sibling(s); and benign hypertension and cerebrovascular accident in the grandparent(s).
OVARNOE02	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from right ovary tissue removed from a 47-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, incisional hernia repair, and panniculectomy. The patient presented with premenopausal menorrhagia. Patient history included osteoarthritis, tubal pregnancy, and polio osteopathy of the left leg. Previous surgeries included gastroenterostomy, plastic repair of the palate, adenotonsillectomy, dilation and curettage, cholecystectomy, and bladder reconstruction. Patient medications included vitamins, iron, and zinc. Family history included benign hypertension and type II diabetes in the father; and type II diabetes in the sibling(s).
OVARTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from left ovary tumor tissue removed from a 44-year-old female. Pathology indicated grade 4 (of 4) serous carcinoma replacing both the right and left ovaries forming solid mass cystic masses. Neoplastic deposits were identified in para-ovarian soft tissue.
SCORNON02	PSPORT1	This normalized spinal cord library was constructed from 3.24M independent clones from the a spinal cord tissue library. RNA was isolated from the spinal cord tissue removed from a 71-year-old Caucasian male who died from respiratory arrest. Patient history included myocardial infarction, gangrene, and end stage renal disease. The normalization and hybridization conditions were adapted from Soares et al.(PNAS (1994) 91:9228).
SINTTMR02	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 59-year-old male. Pathology for the matched tumor tissue indicated multiple (9) carcinoid tumors, grade 1, in the small bowel. The largest tumor was associated with a large mesenteric mass. Multiple convoluted segments of bowel were adhered to the tumor. A single (1 of 13) regional lymph node was positive for malignancy. The peritoneal biopsy indicated focal fat necrosis.

Table 6

Library	Vector	Library Description
THYRNOT03	pINCY	Library was constructed using RNA isolated from thyroid tissue removed from the left thyroid of a 28-year-old Caucasian female during a complete thyroidectomy. Pathology indicated a small nodule of adenomatous hyperplasia present in the left thyroid. Pathology for the associated tumor tissue indicated dominant follicular adenoma, forming a well-encapsulated mass in the left thyroid.
UTRSDIC01	PSPORT1	This large size fractionated library was constructed using pooled cDNA from eight donors. cDNA was generated using mRNA isolated from endometrial tissue removed from a 32-year-old female (donor A); endometrial tissue removed from a 32-year-old Caucasian female (donor B) during abdominal hysterectomy, bilateral salpingo-oophorectomy, and cystocele repair; from diseased endometrium and myometrium tissue removed from a 38-year-old Caucasian female (donor C) during abdominal hysterectomy, bilateral salpingo-oophorectomy, and exploratory laparotomy; from endometrial tissue removed from a 41-year-old Caucasian female (donor D) during abdominal hysterectomy with removal of a solitary ovary; from endometrial tissue removed from a 43-year-old Caucasian female (donor E) during vaginal hysterectomy,
		dilation and curettage, cystocele repair, rectocele repair and cystostomy; and from endometrial tissue removed from a 48-year-old Caucasian female (donor F) during a vaginal hysterectomy, rectocele repair, and bilateral salpingo-oophorectomy. Pathology (A) indicated the endometrium was in secretory phase. Pathology (B) indicated the endometrium was in the proliferative phase. Pathology (C) indicated extensive adenomatous hyperplasia with squamous metaplasia and focal atypia, forming a polypoid mass within the endometrial cavity. The cervix showed chronic cervicitis and squamous metaplasia. Pathology (D, E) indicated the endometrium was secretory phase. Pathology (F) indicated the endometrium was weakly proliferative.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value=1.0E-8 or less *Full Length sequences:* Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs:* fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. & S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART, and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, p. 1-350	*PFAM, INCY, SMART, or TIGRFAM hits:* Probability value=1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

177

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score=120 or greater; Match length=56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
53	7510095	1217456H1	SNP00105562	83	2816	A	A	G	noncoding	n/d	n/d	n/d	n/d
53	7510095	1218806H1	SNP00004978	169	1587	A	G	A	E465	n/a	n/a	n/a	0.97
53	7510095	1286173H1	SNP00096962	43	2180	A	A	G	noncoding	n/d	n/d	n/d	0.97
53	7510095	1351490H1	SNP00139100	226	1770	C	C	T	P526	n/a	n/a	n/a	n/a
53	7510095	1386861H1	SNP00025844	183	936	T	T	G	P248	n/d	n/a	n/a	n/a
53	7510095	1674463H1	SNP00139100	183	1771	C	C	T	P527	n/a	n/a	n/a	n/a
53	7510095	1686583H1	SNP00025845	105	1231	C	C	T	P347	n/a	n/a	n/a	n/a
53	7510095	1798253H1	SNP00004978	194	1584	A	G	A	L464	n/a	n/a	n/a	n/a
53	7510095	1814731H1	SNP00025845	253	1232	C	C	T	P347	n/a	n/a	n/a	n/a
53	7510095	1860842H1	SNP00004978	141	1581	A	G	A	E463	n/d	n/d	n/d	n/d
53	7510095	1908048H1	SNP00105562	132	2817	A	A	G	noncoding	n/d	n/d	n/d	n/d
53	7510095	2061203H1	SNP00105562	34	2810	A	A	G	noncoding	n/d	n/d	n/d	n/d
53	7510095	2109864H1	SNP00096962	84	2179	A	A	G	noncoding	n/d	n/d	n/d	0.97
53	7510095	2500612H1	SNP00025844	178	937	T	T	G	S249	n/d	n/a	n/a	n/a
53	7510095	2600068H1	SNP00004978	80	1583	A	G	A	Q464	n/a	n/a	n/a	n/a
53	7510095	2600068H1	SNP00139100	264	1767	C	C	T	N525	n/a	n/a	n/a	n/a
53	7510095	2745235H1	SNP00025845	170	1229	C	C	T	A346	n/a	n/a	n/a	n/a
53	7510095	2972896H2	SNP00139100	67	1768	C	C	T	P526	n/a	n/a	n/a	n/a
53	7510095	3050947H1	SNP00096962	27	2178	A	A	G	noncoding	n/d	n/d	n/d	0.97
53	7510095	3074463H2	SNP00025845	280	1230	C	C	T	A346	n/a	n/a	n/a	n/a
53	7510095	3246779H1	SNP00025844	186	930	T	T	G	C246	n/d	n/a	n/a	n/a
53	7510095	3336225H1	SNP00004978	36	1585	A	G	A	K465	n/a	n/a	n/a	n/a
53	7510095	3336225H1	SNP00139100	220	1769	C	C	T	P526	n/a	n/a	n/a	n/a
53	7510095	3352432H1	SNP00105562	116	2814	A	A	G	noncoding	n/d	n/d	n/d	n/d
53	7510095	3700090H1	SNP00025845	66	1220	C	C	T	A343	n/a	n/a	n/a	n/a
53	7510095	3831266H1	SNP00025844	69	934	T	T	G	S248	n/d	n/d	n/d	n/a
53	7510095	3859108H1	SNP00004978	64	1586	A	G	A	E465	n/a	n/a	n/a	n/a
53	7510095	3901943H1	SNP00004978	160	1579	G	G	A	E463	n/a	n/a	n/a	n/a
53	7510095	4199364H1	SNP00139100	50	1765	C	C	T	Q525	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
53	7510095	4433120H1	SNP00025844	37	935	G	T	G	R248	n/d	n/a	n/a	n/a
53	7510095	4708619H1	SNP00025845	158	1228	C	C	T	P346	n/a	n/a	n/a	n/a
53	7510095	4858491H1	SNP00004978	78	1582	G	G	A	V464	n/a	n/a	n/a	n/a
53	7510095	4921175H1	SNP00025845	189	1218	C	C	T	S342	n/a	n/a	n/a	n/a
53	7510095	6044386J1	SNP00004979	65	2917	T	C	T	noncoding	n/a	n/a	n/a	n/a
53	7510095	6190082H1	SNP00025844	37	933	T	T	G	H247	n/d	n/a	n/a	n/a
53	7510095	6453815H1	SNP00025844	390	931	T	T	G	Y247	n/d	n/a	n/a	n/a
53	7510095	7972266H1	SNP00130128	232	977	C	T	C	P262	n/a	n/a	n/a	n/a
53	7510095	865177H1	SNP00105562	113	2813	A	A	G	noncoding	n/d	n/d	n/d	n/d
53	7510095	916007H1	SNP00025844	15	929	T	T	G	F246	n/d	n/a	n/a	n/d
54	7510096	1217456H1	SNP00105562	83	2721	A	A	G	noncoding	n/d	n/a	n/a	n/a
54	7510096	1218806H1	SNP00004978	169	1634	A	G	A	noncoding	n/a	n/a	n/a	n/a
54	7510096	1286173H1	SNP00096962	43	2085	A	A	G	noncoding	n/d	n/d	n/d	0.97
54	7510096	1351490H1	SNP00139100	226	1817	C	C	T	noncoding	n/a	n/a	n/a	n/a
54	7510096	1386861H1	SNP00025844	183	980	T	T	G	noncoding	n/d	n/a	n/a	n/a
54	7510096	1674463H1	SNP00139100	183	1818	C	C	T	noncoding	n/a	n/a	n/a	n/a
54	7510096	1686583H1	SNP00025845	105	1278	C	C	T	noncoding	n/a	n/a	n/a	n/a
54	7510096	1798253H1	SNP00004978	194	1631	A	G	A	noncoding	n/a	n/a	n/a	n/a
54	7510096	1814731H1	SNP00025845	253	1279	C	C	A	noncoding	n/a	n/a	n/a	n/a
54	7510096	1860842H1	SNP00004978	141	1628	A	G	A	noncoding	n/a	n/a	n/a	n/a
54	7510096	1908048H1	SNP00105562	132	2722	A	A	G	noncoding	n/d	n/d	n/d	n/d
54	7510096	191849H1	SNP00096962	157	2086	A	A	G	noncoding	n/d	n/d	n/d	0.97
54	7510096	2061203H1	SNP00105562	34	2715	A	A	G	noncoding	n/d	n/d	n/d	n/d
54	7510096	2109864H1	SNP00096962	84	2084	A	A	G	noncoding	n/d	n/d	n/d	0.97
54	7510096	2250586H1	SNP00105562	43	2720	A	A	G	noncoding	n/d	n/d	n/d	n/d
54	7510096	2490326H1	SNP00096962	153	2083	A	A	G	noncoding	n/d	n/a	n/d	0.97
54	7510096	2500612H1	SNP00025844	178	981	T	T	G	noncoding	n/d	n/a	n/a	n/a
54	7510096	2600068H1	SNP00004978	80	1630	A	G	A	noncoding	n/a	n/a	n/a	n/a
54	7510096	2600068H1	SNP00139100	264	1814	C	C	T	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO.	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
54	7510096	2745235H1	SNP00025845	170	1276	C	C	T	noncoding	n/a	n/a	n/a	n/a
54	7510096	2972896H2	SNP00139100	67	1815	C	C	T	noncoding	n/a	n/a	n/a	n/a
54	7510096	3074463H2	SNP00025845	280	1277	C	C	T	noncoding	n/a	n/a	n/a	n/a
54	7510096	3246779H1	SNP00025844	186	974	T	T	G	noncoding	n/d	n/a	n/a	n/a
54	7510096	3336225H1	SNP00004978	36	1632	A	G	A	noncoding	n/a	n/a	n/a	n/a
54	7510096	3336225H1	SNP00139100	220	1816	C	C	T	noncoding	n/a	n/a	n/a	n/a
54	7510096	3352432H1	SNP00105562	116	2719	A	A	G	noncoding	n/d	n/d	n/d	n/d
54	7510096	3700090H1	SNP00025845	66	1267	C	C	T	noncoding	n/a	n/a	n/a	n/a
54	7510096	3831266H1	SNP00025844	69	978	T	T	G	noncoding	n/d	n/a	n/a	n/a
54	7510096	3859108H1	SNP00004978	64	1633	A	G	A	noncoding	n/a	n/a	n/a	n/a
54	7510096	3901943H1	SNP00004978	160	1626	G	G	A	noncoding	n/a	n/a	n/a	n/a
54	7510096	4199364H1	SNP00139100	50	1812	C	C	T	noncoding	n/a	n/a	n/a	n/a
54	7510096	4433120H1	SNP00025844	37	979	G	T	G	noncoding	n/d	n/a	n/a	n/a
54	7510096	4708619H1	SNP00025845	158	1275	C	C	T	noncoding	n/a	n/a	n/a	n/a
54	7510096	4858491H1	SNP00004978	78	1629	G	G	A	noncoding	n/a	n/a	n/a	n/a
54	7510096	4921175H1	SNP00025845	189	1265	C	C	T	noncoding	n/a	n/a	n/a	n/a
54	7510096	5903972H1	SNP00096962	116	2082	A	A	G	noncoding	n/d	n/d	n/d	0.97
54	7510096	6044386J1	SNP00004979	65	2822	T	C	T	noncoding	n/a	n/a	n/a	n/a
54	7510096	6190082H1	SNP00025844	37	977	T	T	G	noncoding	n/d	n/a	n/a	n/a
54	7510096	6453815H1	SNP00025844	390	975	T	T	G	noncoding	n/d	n/d	n/a	n/a
54	7510096	7972266H1	SNP00130128	232	1024	C	T	C	noncoding	n/a	n/a	n/a	n/a
54	7510096	865177H1	SNP00105562	113	2718	A	A	G	noncoding	n/d	n/d	n/d	n/d
54	7510096	916007H1	SNP00025844	15	973	T	T	G	noncoding	n/d	n/a	n/a	n/a

What is claimed is:

    1. An isolated polypeptide selected from the group consisting of:

a)   a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27,

b)   a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2-5, SEQ ID NO:7, SEQ ID NO:9-10, SEQ ID NO:12, SEQ ID NO:14-15, SEQ ID NO:18-19, SEQ ID NO:22-23, and SEQ ID NO:27,

c)   a polypeptide comprising a naturally occurring amino acid sequence at least 98% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1and SEQ ID NO:17,

d)   a polypeptide comprising a naturally occurring amino acid sequence at least 96% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:6 and SEQ ID NO:8,

e)   a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to the amino acid sequence of SEQ ID NO:11,

f)   a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:13 and SEQ ID NO:16,

g)   a polypeptide comprising a naturally occurring amino acid sequence at least 94% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:20, SEQ ID NO:24, and SEQ ID NO:26,

h)   a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and

i)   an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

    2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

    3. An isolated polynucleotide encoding a polypeptide of claim 1.

    4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a
5   polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
10

9. A method of producing a polypeptide of claim 1, the method comprising:

a)   culturing a cell under conditions suitable for expression of the polypeptide, wherein
        said cell is transformed with a recombinant polynucleotide, and said recombinant
        polynucleotide comprises a promoter sequence operably linked to a polynucleotide
15   encoding the polypeptide of claim 1, and

.b)   recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.
20

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a)   a polynucleotide comprising a polynucleotide sequence selected from the group
25   consisting of SEQ ID NO:28-54,

b)   a polynucleotide comprising a naturally occurring polynucleotide sequence at least
        90% identical to a polynucleotide sequence selected from the group consisting of
        SEQ ID NO:28-37, SEQ ID NO:39-51, and SEQ ID NO:54,

c)   a polynucleotide comprising a naturally occurring polynucleotide sequence at least
30   99% identical to a polynucleotide sequence of SEQ ID NO:52,

d)   a polynucleotide comprising a naturally occurring polynucleotide sequence at least
        94% identical to a polynucleotide sequence of SEQ ID NO:54,

e)   a polynucleotide complementary to a polynucleotide of a),

f)   a polynucleotide complementary to a polynucleotide of b),

g)   a polynucleotide complementary to a polynucleotide of c),

h)   a polynucleotide complementary to a polynucleotide of d), and

i)   an RNA equivalent of a)-h).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a
polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide
having a sequence of a polynucleotide of claim 12, the method comprising:

a)   hybridizing the sample with a probe comprising at least 20 contiguous nucleotides
        comprising a sequence complementary to said target polynucleotide in the sample,
        and which probe specifically hybridizes to said target polynucleotide, under
        conditions whereby a hybridization complex is formed between said probe and said
        target polynucleotide or fragments thereof, and

b)   detecting the presence or absence of said hybridization complex, and, optionally, if
        present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide
having a sequence of a polynucleotide of claim 12, the method comprising:

a)   amplifying said target polynucleotide or fragment thereof using polymerase chain
        reaction amplification, and

b)   detecting the presence or absence of said amplified target polynucleotide or fragment
        thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable
excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence
selected from the group consisting of SEQ ID NO:1-27.

19. A method for treating a disease or condition associated with decreased expression of

functional CGDD, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of

5   claim 1, the method comprising:

   a)   exposing a sample comprising a polypeptide of claim 1 to a compound, and

   b)   detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a

10   pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional CGDD, comprising administering to a patient in need of such treatment a composition of claim 21.

15

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

   a)   exposing a sample comprising a polypeptide of claim 1 to a compound, and

   b)   detecting antagonist activity in the sample.

20

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional

25   CGDD, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

   a)   combining the polypeptide of claim 1 with at least one test compound under suitable

30   conditions, and

   b)   detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of

claim 1, the method comprising:

a)   combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b)   assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c)   comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a)   exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b)   detecting altered expression of the target polynucleotide, and

c)   comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a)   treating a biological sample containing nucleic acids with the test compound,

b)   hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c)   quantifying the amount of hybridization complex, and

d)   comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of CGDD in a biological sample, the method comprising:

a)   combining the biological sample with an antibody of claim 11, under conditions

suitable for the antibody to bind the polypeptide and form an antibody:polypeptide

complex, and

b)   detecting the complex, wherein the presence of the complex correlates with the

5   presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) ·   a chimeric antibody,

b)   a single chain antibody,

10   c)   a Fab fragment,

d)   a F(ab')$_2$ fragment, or

e)   a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

15

33. A method of diagnosing a condition or disease associated with the expression of CGDD

in a subject, comprising administering to said subject an effective amount of the composition of claim

32.

20   34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of CGDD

in a subject, comprising administering to said subject an effective amount of the composition of claim

34.

25

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim

11, the method comprising:

a)   immunizing an animal with a polypeptide consisting of an amino acid sequence

selected from the group consisting of SEQ ID NO:1-27, or an immunogenic fragment

30   thereof, under conditions to elicit an antibody response,

b)   isolating antibodies from the animal, and ·

c)   screening the isolated antibodies with the polypeptide, thereby identifying a

polyclonal antibody which specifically binds to a polypeptide comprising an amino

acid sequence selected from the group consisting of SEQ ID NO:1-27.

35

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

5   39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a)   immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

10   b)   isolating antibody producing cells from the animal,

c)   fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d)   culturing the hybridoma cells, and

e)   isolating from the culture monoclonal antibody which specifically binds to a

15   polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

40. A monoclonal antibody produced by a method of claim 39.

20   41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

25   43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27 in a sample, the method comprising:

30   a)   incubating the antibody of claim 11 with a sample under conditions to allow specific .
               binding of the antibody and the polypeptide, and

b)   detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27 in the sample.

35

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27 from a sample, the method comprising:

    a)   incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

5   b)   separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim

10   13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

    a)   labeling the polynucleotides of the sample,

15   b)   contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

    c)   quantifying the expression of the polynucleotides in the sample.

20   48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

25

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is

30   completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

35   52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to 5 said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains 10 nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

15   57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

20

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

25   62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

30

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

35   67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

5   70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

10

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

15   75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

20

78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.

79. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.

25   80. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.

81. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.

82. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:27.

30

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
35   NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
5   NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

10   88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

15

90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
20   NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

25   93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

30

95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
35   NO:41.

97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
5   NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.

10   100. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:45.

101. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:46.

15

102. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:47.

103. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
20   NO:48.

104. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:49.

25   105. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:50.

106. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:51.

30

107. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:52.

108. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
35   NO:53.

109. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:54.

<110> INCYTE GENOMICS, INC.
      AZIMZAI, Yalda
      BARROSO, Ines
      BAUGHN, Mariah R.
      BECHA, Shanya D.
      BOROWSKY, Mark L.
      DUGGAN, Brendan M.
      ELLIOTT, Vicki S.
      EMERLING, Brooke M.
      FORSYTHE, Ian J.
      GIETZEN, Kimberly J.
      GORVAD, Ann E.
      GRAUL, Richard C.
      GRIFFIN, Jennifer A.
      GURURAJAN, Rajagopal
      HAFALIA, April J.A.
      ISON, Craig H.
      KABLE, Amy E.
      KHAN, Farrah A.
      LEE, Sally
      LEE, Soo Yeun
      LI, Joana X.
      REDDY, Roopa
      RICHARDON, Thomas W.
      SPRAGUE, William W.
      SWARNAKAR, Anita
      TANG, Y. Tom
      WARREN, Bridget A.
      XU, Yuming
      YAO, Monique G.
      YUE, Henry
      YUE, Huibin

<120> PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH

<130> PF-1126 PCT

<140> To Be Assigned
<141> Herewith

<150> US 60/311,017
<151> 2001-08-08

<150> US 60/313,070
<151> 2001-08-17

<150> US 60/313,071
<151> 2001-08-17

<150> US 60/314,678
<151> 2001-08-24

<150> US 60/316,692
<151> 2001-08-31

<150> US 60/317,913
<151> 2001-09-07

<150> US 60/322,182
<151> 2001-09-14

<150> US 60/340,747
<151> 2001-12-07

<150> US 60/342,761
<151> 2001-12-20

<150> US 60/369,129
<151> 2002-03-29

<160> 54

<170> PERL Program

<210> 1
<211> 168
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 261510CD1

<400> 1
Met Glu Pro Pro Asn Leu Tyr Pro Val Lys Leu Tyr Val Tyr Asp
1   5   10   15
Leu Ser Lys Gly Leu Ala Arg Arg Leu Ser Pro Ile Met Leu Gly
                20   25   30
Lys Gln Leu Glu Gly Ile Trp His Thr Ser Ile Val Val His Lys
                35   40   45
Asp Glu Phe Phe Phe Gly Ser Gly Gly Ile Ser Ser Cys Pro Pro
                    50   55   60
Gly Gly Thr Leu Leu Gly Pro Pro Asp Ser Val Val Asp Val Gly
                65   70   75
Ser Thr Glu Val Thr Glu Glu Ile Phe Leu Glu Tyr Leu Ser Ser
                80   85   90
Leu Gly Glu Ser Leu Phe Arg Gly Glu Ala Tyr Asn Leu Phe Glu
                95   100   105
His Asn Cys Asn Thr Phe Ser Asn Glu Val Ala Gln Phe Leu Thr
                110   115   120
Gly Arg Lys Ile Pro Ser Tyr Ile Thr Asp Leu Pro Ser Glu Val
                125   130   135
Leu Ser Thr Pro Phe Gly Gln Ala Leu Arg Pro Leu Leu Asp Ser
                140   145   150
Ile Gln Ile Gln Pro Pro Gly Gly Ser Ser Val Gly Arg Pro Asn
                155   160   165
Gly Gln Ser

<210> 2
<211> 548
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7498674CD1

<400> 2
Met Ala Gly Leu Arg Ala Arg Gly Gly Pro Gly Pro Gly Leu Leu
1   5   10   15
Ala Leu Ser Ala Leu Gly Phe Cys Leu Met Leu Gln Val Ser Ala
                20   25   30
Lys Arg Pro Pro Lys Thr Pro Pro Cys Pro Pro Ser Cys Ser Cys
                35   40   45
Thr Arg Asp Thr Ala Phe Cys Val Asp Ser Lys Ala Val Pro Arg
                    50   55   60

```
Asn Leu Pro Ser Glu Val Ile Ser Leu Thr Leu Val Asn Ala Ala
             65                  70                  75
Phe Ser Glu Ile Gln Asp Gly Ala Phe Ser His Leu Pro Leu Leu
             80                  85                  90
Gln Phe Leu Leu Leu Asn Ser Asn Lys Phe Thr Leu Ile Gly Asp
             95                 100                 105
Asn Ala Phe Thr Gly Leu Ser His Leu Gln Tyr Leu Phe Ile Glu
            110                 115                 120
Asn Asn Asp Ile Trp Ala Leu Ser Lys Phe Thr Phe Arg Gly Leu
            125                 130                 135
Lys Ser Leu Thr His Leu Ser Leu Ala Asn Asn Asn Leu Gln Thr
            140                 145                 150
Leu Pro Arg Asp Ile Phe Arg Pro Leu Asp Ile Leu Asn Asp Leu
            155                 160                 165
Asp Leu Arg Gly Asn Ser Leu Asn Cys Asp Cys Lys Val Lys Trp
            170                 175                 180
Leu Val Glu Trp Leu Ala His Thr Asn Thr Thr Val Ala Pro Ile
            185                 190                 195
Tyr Cys Ala Ser Pro Pro Arg Phe Gln Glu His Lys Val Gln Asp
            200                 205                 210
Leu Pro Leu Arg Glu Phe Asp Cys Ile Thr Thr Asp Phe Val Leu
            215                 220                 225
Tyr Gln Thr Leu Ala Phe Pro Ala Val Ser Ala Glu Pro Phe Leu
            230                 235                 240
Tyr Ser Ser Asp Leu Tyr Leu Ala Leu Ala Gln Pro Gly Val Ser
            245                 250                 255
Ala Cys Thr Ile Leu Lys Trp Asp Tyr Val Glu Arg Gln Leu Arg
            260                 265                 270
Asp Tyr Asp Arg Ile Pro Ala Pro Ser Ala Val His Cys Lys Pro
            275                 280                 285
Met Val Val Asp Ser Gln Leu Tyr Val Val Val Ala Gln Leu Phe
            290                 295                 300
Gly Gly Ser Tyr Ile Tyr His Trp Asp Pro Asn Thr Thr Arg Phe
            305                 310                 315
Thr Arg Leu Gln Asp Ile Asp Pro Gln Arg Val Arg Lys Pro Asn
            320                 325                 330
Asp Leu Glu Ala Phe Arg Ile Asp Gly Asp Trp Tyr Phe Ala Val
            335                 340                 345
Ala Asp Ser Ser Lys Ala Gly Ala Thr Ser Leu Tyr Arg Trp His
            350                 355                 360
Gln Asn Gly Phe Tyr Ser His Gln Ala Leu His Pro Trp His Arg
            365                 370                 375
Asp Thr Asp Leu Glu Phe Val Asp Gly Glu Gly Lys Pro Arg Leu
            380                 385                 390
Ile Val Ser Ser Ser Ser Gln Ala Pro Val Ile Tyr Gln Trp Ser
            395                 400                 405
Arg Thr Gln Lys Gln Phe Val Ala Gln Gly Glu Val Thr Gln Val
            410                 415                 420
Pro Asp Ala Gln Ala Val Lys His Phe Arg Ala Gly Arg Asp Ser
            425                 430                 435
Tyr Leu Cys Leu Ser Arg Tyr Ile Gly Asp Ser Lys Ile Leu Arg
            440                 445                 450
Trp Glu Gly Thr Arg Phe Ser Glu Val Gln Ala Leu Pro Ser Arg
            455                 460                 465
Gly Ser Leu Ala Leu Gln Pro Phe Leu Val Gly Gly Arg Arg Tyr
            470                 475                 480
Leu Ala Leu Gly Ser Asp Phe Ser Phe Thr Gln Ile Tyr Gln Trp
            485                 490                 495
Asp Glu Gly Arg Gln Lys Phe Val Arg Phe Gln Glu Leu Ala Val
            500                 505                 510
Gln Ala Pro Arg Ala Phe Cys Tyr Met Pro Ala Gly Asp Ala Gln
            515                 520                 525
Leu Leu Leu Ala Pro Ser Phe Lys Gly Gln Thr Leu Val Tyr Arg
```

```
                  530                    535                    540
His Ile Val Val Asp Leu Ser Ala
                  545


<210> 3
<211> 378
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1629617CD1

<400> 3
Met Thr Gly Ala Ile Ser Tyr His Pro Tyr Gly Ser Ala Ala Tyr
  1               5                  10                  15
Pro Tyr Gln Leu Asn Asp Pro Ala Tyr Arg Lys Asn Ala Thr Arg
                  20                  25                  30
Asp Ala Thr Ala Thr Leu Lys Ala Trp Leu Asn Glu His Arg Lys
                  35                  40                  45
Asn Pro Tyr Pro Thr Lys Gly Glu Lys Ile Met Leu Ala Ile Ile
                  50                  55                  60
Thr Lys Met Thr Leu Thr Gln Val Ser Thr Trp Phe Ala Asn Ala
                  65                  70                  75
Arg Arg Arg Leu Lys Lys Glu Asn Lys Met Thr Trp Ala Pro Arg
                  80                  85                  90
Asn Lys Ser Glu Asp Glu Asp Glu Asp Glu Gly Asp Ala Thr Arg
                  95                  100                 105
Ser Lys Asp Glu Ser Pro Asp Lys Ala Gln Glu Gly Thr Glu Thr
                  110                 115                 120
Ser Ala Glu Asp Glu Gly Ile Ser Leu His Val Asp Ser Leu Thr
                  125                 130                 135
Asp His Ser Cys Ser Ala Glu Ser Asp Gly Glu Lys Leu Pro Cys
                  140                 145                 150
Arg Ala Gly Asp Pro Leu Cys Glu Ser Gly Ser Glu Cys Lys Asp
                  155                 160                 165
Lys Tyr Asp Asp Leu Glu Asp Asp Glu Asp Asp Glu Glu Gly
                  170                 175                 180
Glu Arg Gly Leu Ala Pro Pro Lys Pro Val Thr Ser Ser Pro Leu
                  185                 190                 195
Thr Gly Leu Glu Ala Pro Leu Leu Ser Pro Pro Glu Ala Ala
                  200                 205                 210
Pro Arg Gly Gly Arg Lys Thr Pro Gln Gly Ser Arg Thr Ser Pro
                  215                 220                 225
Gly Ala Pro Pro Ala Ser Lys Pro Lys Leu Trp Ser Leu Ala
                  230                 235                 240
Glu Ile Ala Thr Ser Asp Leu Lys Gln Pro Ser Leu Gly Pro Gly
                  245                 250                 255
Cys Gly Pro Pro Gly Leu Pro Ala Ala Ala Pro Ala Ser Thr
                  260                 265                 270
Gly Ala Pro Pro Gly Gly Ser Pro Tyr Pro Ala Ser Pro Leu Leu
                  275                 280                 285
Gly Arg Pro Leu Tyr Tyr Thr Ser Pro Phe Tyr Gly Asn Tyr Thr
                  290                 295                 300
Asn Tyr Gly Asn Leu Asn Ala Ala Leu Gln Gly Gln Gly Leu Leu
                  305                 310                 315
Arg Tyr Asn Ser Ala Ala Ala Ala Pro Gly Glu Ala Leu His Thr
                  320                 325                 330
Ala Pro Lys Ala Ala Ser Asp Ala Gly Lys Ala Gly Ala His Pro
                  335                 340                 345
Leu Glu Ser His Tyr Arg Ser Pro Gly Gly Gly Tyr Glu Pro Lys
                  350                 355                 360
Lys Asp Ala Ser Glu Gly Cys Thr Val Val Gly Gly Gly Val Gln
```

```
                    365                 370                 375
Pro Tyr Leu


<210> 4
<211> 528
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2369279CD1

<400> 4
Met Glu Val Val Pro Ala Glu Ile Ala Lys Lys Pro Arg Thr Pro
 1               5                  10                  15
Thr Ser Gly Pro Val Ile Thr Lys Leu Ile Phe Ala Lys Pro Ile
                20                  25                  30
Asn Ser Lys Ala Val Thr Gly Gln Thr Thr Gln Val Ser Pro Pro
                35                  40                  45
Val Ile Ala Gly Arg Val Leu Ser Gln Ser Thr Pro Gly Thr Pro
                50                  55                  60
Ser Lys Thr Ile Thr Ile Ser Glu Ser Gly Val Ile Gly Ser Thr
                65                  70                  75
Leu Asn Ser Thr Thr Gln Thr Pro Asn Lys Ile Ala Ile Ser Pro
                80                  85                  90
Leu Lys Ser Pro Asn Lys Ala Val Lys Ser Thr Val Gln Thr Ile
                95                  100                 105
Thr Val Gly Gly Val Ser Thr Ser Gln Phe Lys Thr Ile Ile Pro
                110                 115                 120
Leu Ala Thr Ala Pro Asn Val Gln Gln Ile Gln Val Pro Gly Ser
                125                 130                 135
Lys Phe His Tyr Val Arg Leu Val Thr Ala Thr Ser Ala Ser Ser
                140                 145                 150
Ser Thr Gln Pro Val Ser Gln Asn Pro Ser Thr Asn Thr Gln Pro
                155                 160                 165
Leu Gln Gln Ala Lys Pro Val Val Val Asn Thr Thr Pro Val Arg
                170                 175                 180
Met Ser Val Pro Ile Val Ser Ala Gln Ala Val Lys Gln Val Val
                185                 190                 195
Pro Lys Pro Ile Asn Pro Thr Ser Gln Ile Val Thr Thr Ser Gln
                200                 205                 210
Pro Gln Gln Arg Leu Ile Met Pro Ala Thr Pro Leu Pro Gln Ile
                215                 220                 225
Gln Pro Asn Leu Thr Asn Leu Pro Pro Gly Thr Val Leu Ala Pro
                230                 235                 240
Ala Pro Gly Thr Gly Asn Val Gly Tyr Ala Val Leu Pro Ala Gln
                245                 250                 255
Tyr Val Thr Gln Leu Gln Gln Ser Ser Tyr Val Ser Ile Ala Ser
                260                 265                 270
Asn Ser Thr Phe Thr Gly Thr Ser Gly Ile Gln Thr Gln Ala Arg
                275                 280                 285
Leu Pro Phe Asn Gly Ile Ile Pro Ser Glu Ser Ala Ser Arg Pro
                290                 295                 300
Arg Lys Pro Cys Asn Cys Thr Lys Ser Leu Cys Leu Lys Leu Tyr
                305                 310                 315
Cys Asp Cys Phe Ala Asn Gly Glu Phe Cys Asn Asn Cys Asn Cys
                320                 325                 330
Thr Asn Cys Tyr Asn Asn Leu Glu His Glu Asn Glu Arg Gln Lys
                335                 340                 345
Ala Ile Lys Ala Cys Leu Asp Arg Asn Pro Glu Ala Phe Lys Pro
                350                 355                 360
Lys Ile Gly Lys Gly Lys Glu Gly Glu Ser Asp Arg Arg His Ser
```

```
                365                    370                    375
Lys Gly Cys Asn Cys Lys Arg Ser Gly Cys Leu Lys Asn Tyr Cys
            380                    385                    390
Glu Cys Tyr Glu Ala Lys Ile Met Cys Ser Ser Ile Cys Lys Cys
            395                    400                    405
Ile Gly Cys Lys Asn Phe Glu Glu Ser Pro Glu Arg Lys Thr Leu
            410                    415                    420
Met His Leu Ala Asp Ala Ala Glu Val Arg Val Gln Gln Gln Thr
            425                    430                    435
Ala Ala Lys Thr Lys Leu Ser Ser Gln Ile Ser Asp Leu Leu Thr
            440                    445                    450
Arg Pro Thr Pro Ala Leu Asn Ser Gly Gly Gly Lys Leu Pro Phe
            455                    460                    465
Thr Phe Val Thr Lys Glu Val Ala Glu Ala Thr Cys Asn Cys Leu
            470                    475                    480
Leu Ala Gln Ala Glu Gln Ala Asp Lys Lys Gly Lys Ser Lys Ala
            485                    490                    495
Ala Ala Glu Arg Met Ile Leu Glu Glu Phe Gly Arg Cys Leu Met
            500        ·           505                    510
Ser Val Ile Asn Ser Ala Gly Lys Ala Lys Ser Asp Pro Cys Ala
            515                    520                    525
Met Asn Cys
```

```
<210> 5
<211> 374
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3167506CD1

<400> 5
Met Ala Ala Pro Pro Arg Pro Ala Pro Ser Pro Pro Ala Pro Arg
 1               5                  10                  15
Arg Leu Asp Thr Ser Asp Val Leu Gln Gln Ile Met Ala Ile Thr
            20                  25                  30
Asp Gln Ser Leu Asp Glu Ala Gln Ala Arg Lys His Ala Leu Asn
            35                  40                  45
Cys His Arg Met Lys Pro Ala Leu Phe Ser Val Leu Cys Glu Ile
            50                  55                  60
Lys Glu Lys Thr Val Val Ser Ile Arg Gly Ile Gln Asp Glu Asp
            65                  70                  75
Pro Pro Asp Ala Gln Leu Leu Arg Leu Asp Asn Met Leu Leu Ala
            80                  85                  90
Glu Gly Val Cys Arg Pro Glu Lys Arg Gly Arg Gly Gly Ala Val
            95                  100                 105
Ala Arg Ala Gly Thr Ala Thr Pro Gly Gly Cys Pro Asn Asp Asn
            110                 115                 120
Ser Ile Glu His Ser Asp Tyr Arg Ala Lys Leu Ser Gln Ile Arg
            125                 130                 135
Gln Ile Tyr His Ser Glu Leu Glu Lys Tyr Glu Gln Ala Cys Arg
            140                 145                 150
Glu Phe Thr Thr His Val Thr Asn Leu Leu Gln Glu Gln Ser Arg
            155                 160                 165
Met Arg Pro Val Ser Pro Lys Glu Ile Glu Arg Met Val Gly Ala
            170                 175                 180
Ile His Gly Lys Phe Ser Ala Ile Gln Met Gln Leu Lys Gln Ser
            185                 190                 195
Thr Cys Glu Ala Val Met Thr Leu Arg Ser Arg Leu Leu Asp Ala
            200                 205                 210
Arg Arg Lys Arg Arg Asn Phe Ser Lys Gln Ala Thr Glu Val Leu
```

6/71

```
                215                  220                  225
Asn Glu Tyr Phe Tyr Ser His Leu Asn Asn Pro Tyr Pro Ser Glu
            230                  235                  240
Glu Ala Lys Glu Glu Leu Ala Arg Lys Gly Gly Leu Thr Ile Ser
            245                  250                  255
Gln Val Ser Asn Trp Phe Gly Asn Lys Arg Ile Arg Tyr Lys Lys
            260                  265                  270
Asn Met Gly Lys Phe Gln Glu Glu Ala Thr Ile Tyr Thr Gly Lys
            275                  280                  285
Thr Ala Val Asp Thr Thr Glu Val Gly Val Pro Gly Asn His Ala
            290                  295                  300
Ser Cys Leu Ser Thr Pro Ser Ser Gly Ser Ser Gly Pro Phe Pro
            305                  310                  315
Leu Pro Ser Ala Gly Asp Ala Phe Leu Thr Leu Arg Thr Leu Ala
            320                  325                  330
Ser Leu Gln Pro Pro Pro Gly Gly Gly Cys Leu Gln Ser Gln Ala
            335                  340                  345
Gln Gly Ser Trp Gln Gly Ala Thr Pro Gln Pro Ala Thr Ala Ser
            350                  355                  360
Pro Ala Gly Asp Pro Gly Ser Ile Asn Ser Ser Thr Ser Asn
            365                  370
```

```
<210> 6
<211> 1723
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3075937CD1

<400> 6
Met Glu Lys Arg Ser Ser Gly Arg Arg Ser Gly Arg Arg Arg Gly
 1               5                  10                  15
Ser Gln Lys Ser Thr Asp Ser Pro Gly Ala Asp Ala Glu Leu Pro
                20                  25                  30
Glu Ser Ala Ala Arg Asp Asp Ala Val Phe Asp Asp Glu Val Ala
                35                  40                  45
Pro Asn Ala Ala Ser Asp Asn Ala Ser Ala Glu Lys Lys Val Lys
                50                  55                  60
Ser Pro Arg Ala Ala Leu Asp Gly Gly Val Ala Ser Ala Ala Ser
                65                  70                  75
Pro Glu Ser Lys Pro Ser Pro Gly Thr Lys Gly Gln Leu Arg Gly
                80                  85                  90
Glu Ser Asp Arg Ser Lys Gln Pro Pro Pro Ala Ser Ser Pro Thr
                95                  100                 105
Lys Arg Lys Gly Arg Ser Arg Ala Leu Glu Ala Val Pro Ala Pro
                110                 115                 120
Pro Ala Ser Gly Pro Arg Ala Pro Ala Lys Glu Ser Pro Pro Lys
                125                 130                 135
Arg Val Pro Asp Pro Ser Pro Val Thr Lys Gly Thr Ala Ala Glu
                140                 145                 150
Ser Gly Glu Glu Ala Ala Arg Ala Ile Ser Arg Glu Leu Pro Val
                155                 160                 165
Lys Ser Ser Ser Leu Leu Pro Glu Ile Lys Pro Glu His Lys Arg
                170                 175                 180
Gly Pro Leu Pro Asn His Phe Asn Gly Arg Ala Glu Gly Gly Arg
                185                 190                 195
Ser Arg Glu Leu Gly Arg Ala Ala Gly Ala Pro Gly Ala Ser Asp
                200                 205                 210
Ala Asp Gly Leu Lys Pro Arg Asn His Phe Gly Val Gly Arg Ser
                215                 220                 225
Thr Val Thr Thr Lys Val Thr Leu Pro Ala Lys Pro Lys His Val
```

```
                    230                   235                   240
Glu Leu Asn Leu Lys Thr Pro Lys Asn Leu Asp Ser Leu Gly Asn
                    245                   250                   255
Glu His Asn Pro Phe Ser Gln Pro Val His Lys Gly Asn Thr Ala
                    260                   265                   270
Thr Lys Ile Ser Leu Phe Glu Asn Lys Arg Thr Asn Ser Ser Pro
                    275                   280                   285
Arg His Thr Asp Ile Arg Gly Pro Arg Asn Thr Pro Ala Ser Ser
                    290                   295                   300
Lys Thr Phe Val Gly Arg Ala Lys Leu Asn Leu Ala Lys Lys Ala
                    305                   310                   315
Lys Glu Met Glu Gln Pro Glu Lys Lys Val Met Pro Asn Ser Pro
                    320                   325                   330
Gln Asn Gly Val Leu Val Lys Glu Thr Ala Ile Glu Thr Lys Val
                    335                   340                   345
Thr Val Ser Glu Glu Glu Ile Leu Pro Ala Thr Arg Gly Met Asn
                    350                   355                   360
Gly Asp Ser Ser Glu Asn Gln Ala Leu Gly Pro Gln Pro Asn Gln
                    365                   370                   375
Asp Asp Lys Ala Asp Val Gln Thr Asp Ala Gly Cys Leu Ser Glu
                    380                   385                   390
Pro Val Ala Ser Ala Leu Ile Pro Val Lys Asp His Lys Leu Leu
                    395                   400                   405
Glu Lys Glu Asp Ser Glu Ala Ala Asp Ser Lys Ser Leu Val Leu
                    410                   415                   420
Glu Asn Val Thr Asp Thr Ala Gln Asp Ile Pro Thr Thr Val Asp
                    425                   430                   435
Thr Lys Asp Leu Pro Pro Thr Ala Met Pro Lys Pro Gln His Thr
                    440                   445                   450
Phe Ser Asp Ser Gln Ser Pro Ala Glu Ser Ser Pro Gly Pro Ser
                    455                   460                   465
Leu Ser Leu Ser Ala Pro Ala Pro Gly Asp Val Pro Lys Asp Thr
                    470                   475                   480
Cys Val Gln Ser Pro Ile Ser Ser Phe Pro Cys Thr Asp Leu Lys
                    485                   490                   495
Val Ser Glu Asn His Lys Gly Cys Val Leu Pro Val Ser Arg Gln
                    500                   505                   510
Asn Asn Glu Lys Met Pro Leu Leu Glu Leu Gly Gly Glu Thr Thr
                    515                   520                   525
Pro Pro Leu Ser Thr Glu Arg Ser Pro Glu Ala Val Gly Ser Glu
                    530                   535                   540
Cys Pro Ser Arg Val Leu Val Gln Val Arg Ser Phe Val Leu Pro
                    545                   550                   555
Val Glu Ser Thr Gln Asp Val Ser Ser Gln Val Ile Pro Glu Ser
                    560                   565                   570
Ser Glu Val Arg Glu Val Gln Leu Pro Thr Cys His Ser Asn Glu
                    575                   580                   585
Pro Glu Val Val Ser Val Ala Ser Cys Ala Pro Pro Gln Glu Glu
                    590                   595                   600
Val Leu Gly Asn Glu His Ser His Cys Thr Ala Glu Leu Ala Ala
                    605                   610                   615
Lys Ser Gly Pro Gln Val Ile Pro Pro Ala Ser Glu Lys Thr Leu
                    620                   625                   630
Pro Ile Gln Ala Gln Ser Gln Gly Ser Arg Thr Pro Leu Met Ala
                    635                   640                   645
Glu Ser Ser Pro Thr Asn Ser Pro Ser Ser Gly Asn His Leu Ala
                    650                   655                   660
Thr Pro Gln Arg Pro Asp Gln Thr Val Thr Asn Gly Gln Asp Ser
                    665                   670                   675
Pro Ala Ser Leu Leu Asn Ile Ser Ala Gly Ser Asp Asp Ser Val
                    680                   685                   690
Phe Asp Ser Ser Ser Asp Met Glu Lys Phe Thr Glu Ile Ile Lys
                    695                   700                   705
```

Gln Met Asp Ser Ala Val Cys Met Pro Met Lys Arg Lys Lys Ala
            710   715   720
Arg Met Pro Asn Ser Pro Ala Pro His Phe Ala Met Pro Pro Ile
            725   730   735
His Glu Asp His Leu Glu Lys Val Phe Asp Pro Lys Val Phe Thr
            740   745   750
Phe Gly Leu Gly Lys Lys Lys Glu Ser Gln Pro Glu Met Ser Pro
            755   760   765
Ala Leu His Leu Met Gln Asn Leu Asp Thr Lys Ser Lys Leu Arg
            770   775   780
Pro Lys Arg Ala Ser Ala Glu Gln Ser Val Leu Phe Lys Ser Leu
            785   790   795
His Thr Asn Thr Asn Gly Asn Ser Glu Pro Leu Val Met Pro Glu
            800   805   810
Ile Asn Asp Lys Glu Asn Arg Asp Val Thr Asn Gly Gly Ile Lys
            815   820   825
Arg Ser Arg Leu Glu Lys Ser Ala Leu Phe Ser Ser Leu Leu Ser
            830   835   840
Ser Leu Pro Gln Asp Lys Ile Phe Ser Pro Ser Val Thr Ser Val
            845   850   855
Asn Thr Met Thr Thr Ala Phe Ser Thr Ser Gln Asn Gly Ser Leu
            860   865   870
Ser Gln Ser Ser Val Ser Gln Pro Thr Thr Glu Gly Ala Pro Pro
            875   880   885
Cys Gly Leu Asn Lys Glu Gln Ser Asn Leu Leu Pro Asp Asn Ser
            890   895   900
Leu Lys Val Phe Asn Phe Asn Ser Ser Ser Thr Ser His Ser Ser
            905   910   915
Leu Lys Ser Pro Ser His Met Glu Lys Tyr Pro Gln Lys Glu Lys
            920   925   930
Thr Lys Glu Asp Leu Asp Ser Arg Ser Asn Leu His Leu Pro Glu
            935   940   945
Thr Lys Phe Ser Glu Leu Ser Lys Leu Lys Asn Asp Asp Met Glu
            950   955   960
Lys Ala Asn His Ile Glu Ser Val Ile Lys Ser Asn Leu Pro Asn
            965   970   975
Cys Ala Asn Ser Asp Thr Asp Phe Met Gly Leu Phe Lys Ser Ser
            980   985   990
Arg Tyr Asp Pro Ser Ile Ser Phe Ser Gly Met Ser Leu Ser Asp
            995   1000   1005
Thr Met Thr Leu Arg Gly Ser Val Gln Asn Lys Leu Asn Pro Arg
            1010   1015   1020
Pro Gly Lys Val Val Ile Tyr Ser Glu Pro Asp Val Ser Glu Lys
            1025   1030   1035
Cys Ile Glu Val Phe Ser Asp Ile Gln Asp Cys Ser Ser Trp Ser
            1040   1045   1050
Leu Ser Pro Val Ile Leu Ile Lys Val Val Arg Gly Cys Trp Ile
            1055   1060   1065
Leu Tyr Glu Gln Pro Asn Phe Glu Gly His Ser Ile Pro Leu Glu
            1070   1075   1080
Glu Gly Glu Leu Glu Leu Ser Gly Leu Trp Gly Ile Glu Asp Ile
            1085   1090   1095
Leu Glu Arg His Glu Glu Ala Glu Ser Asp Lys Pro Val Val Ile
            1100   1105   1110
Gly Ser Ile Arg His Val Val Gln Asp Tyr Arg Val Ser His Ile
            1115   1120   1125
Asp Leu Phe Thr Glu Pro Glu Gly Leu Gly Ile Leu Ser Ser Tyr
            1130   1135   1140
Phe Asp Asp Thr Glu Glu Met Gln Gly Phe Gly Val Met Gln Lys
            1145   1150   1155
Thr Cys Ser Met Lys Val His Trp Gly Thr Trp Leu Ile Tyr Glu
            1160   1165   1170
Glu Pro Gly Phe Gln Gly Val Pro Phe Ile Leu Glu Pro Gly Glu

```
                 1175                   1180                   1185
Tyr Pro Asp Leu Ser Phe Trp Asp Thr Glu Ala Ala Tyr Ile Gly
                 1190                   1195                   1200
Ser Met Arg Pro Leu Lys Met Gly Gly Arg Lys Val Glu Phe Pro
                 1205                   1210                   1215
Thr Asp Pro Lys Val Val Val Tyr Glu Lys Pro Phe Phe Glu Gly
                 1220                   1225                   1230
Lys Cys Val Glu Leu Glu Thr Gly Met Cys Ser Phe Val Met Glu
                 1235                   1240                   1245
Gly Gly Glu Thr Glu Glu Ala Thr Gly Asp Asp His Leu Pro Phe
                 1250                   1255                   1260
Thr Ser Val Gly Ser Met Lys Val Leu Arg Gly Ile Trp Val Ala
                 1265                   1270                   1275
Tyr Glu Lys Pro Gly Phe Thr Gly His Gln Tyr Leu Leu Glu Glu
                 1280                   1285                   1290
Gly Glu Tyr Arg Asp Trp Lys Ala Trp Gly Gly Tyr Asn Gly Glu
                 1295                   1300                   1305
Leu Gln Ser Leu Arg Pro Ile Leu Gly Asp Phe Ser Asn Ala His
                 1310                   1315                   1320
Met Ile Met Tyr Ser Glu Lys Asn Phe Gly Ser Lys Gly Ser Ser
                 1325                   1330                   1335
Ile Asp Val Leu Gly Ile Val Ala Asn Leu Lys Glu Thr Gly Tyr
                 1340                   1345                   1350
Gly Val Lys Thr Gln Ser Ile Asn Val Leu Ser Gly Val Trp Val
                 1355                   1360                   1365
Ala Tyr Glu Asn Pro Asp Phe Thr Gly Glu Gln Tyr Ile Leu Asp
                 1370                   1375                   1380
Lys Gly Phe Tyr Thr Ser Phe Glu Asp Trp Gly Gly Lys Asn Cys
                 1385                   1390                   1395
Lys Ile Ser Ser Val Gln Pro Ile Cys Leu Asp Ser Phe Thr Gly
                 1400                   1405                   1410
Pro Arg Arg Arg Asn Gln Ile His Leu Phe Ser Glu Pro Gln Phe
                 1415                   1420        .          1425
Gln Gly His Ser Gln Ser Phe Glu Glu Thr Thr Ser Gln Ile Asp
                 1430                   1435                   1440
Asp Ser Phe Ser Thr Lys Ser Cys Arg Val Ser Gly Gly Ser Trp
                 1445                   1450        .          1455
Val Val Tyr Asp Gly Glu Asn Phe Thr Gly Asn Gln Tyr Val Leu
                 1460                   1465                   1470
Glu Glu Gly His Tyr Pro Cys Leu Ser Ala Met Gly Cys Pro Pro
                 1475                   1480                   1485
Gly Ala Thr Phe Lys Ser Leu Arg Phe Ile Asp Val Glu Phe Ser
                 1490                   1495                   1500
Glu Pro Thr Ile Ile Leu Phe Glu Arg Glu Asp Phe Lys Gly Lys
                 1505                   1510                   1515
Lys Ile Glu Leu Asn Ala Glu Thr Val Asn Leu Arg Ser Leu Gly
                 1520                   1525                   1530
Phe Asn Thr Gln Ile Arg Ser Val Gln Val Ile Gly Gly Ile Trp
                 1535                   1540                   1545
Val Thr Tyr Glu Tyr Gly Ser Tyr Arg Gly Arg Gln Phe Leu Leu
                 1550                   1555                   1560
Ser Pro Ala Glu Val Pro Asn Trp Tyr Glu Phe Ser Gly Cys Arg
                 1565                   1570                   1575
Gln Ile Gly Ser Leu Arg Pro Phe Val Gln Lys Arg Ile Tyr Phe
                 1580                   1585                   1590
Arg Leu Arg Asn Lys Ala Thr Gly Leu Phe Met Ser Thr Asn Gly
                 1595                   1600                   1605
Asn Leu Glu Asp Leu Lys Leu Leu Arg Ile Gln Val Met Glu Asp
                 1610                   1615                   1620
Val Gly Ala Asp Gln Ile Trp Ile Tyr Gln Glu Gly Cys Ile
                 1625                   1630                   1635
Lys Cys Arg Ile Ala Glu Asp Cys Cys Leu Thr Ile Val Gly Ser
                 1640                   1645                   1650
```

```
Leu Val Thr Ser Gly Ser Lys Leu Gly Leu Ala Leu Asp Gln Asn
         1655             1660                 1665
Ala Asp Ser Gln Phe Trp Ser Leu Lys Ser Asp Gly Arg Ile Tyr
         1670             1675                 1680
Ser Lys Leu Lys Pro Asn Leu Val Leu Asp Ile Lys Gly Gly Thr
         1685             1690                 1695
Gln Tyr Asp Gln Asn His Ile Ile Leu Asn Thr Val Ser Lys Glu
         1700             1705                 1710
Lys Phe Thr Gln Val Trp Glu Ala Met Val Leu Tyr Thr
         1715             1720
```

<210> 7
<211> 255
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5176268CD1

<400> 7
```
Met Arg Arg Ile Ser Leu Thr Ser Ser Pro Val Arg Leu Leu Leu
  1           5               10                  15
Phe Leu Leu Leu Leu Leu Ile Ala Leu Glu Ile Met Val Gly Gly
         20                  25                  30
His Ser Leu Cys Phe Asn Phe Thr Ile Lys Ser Leu Ser Arg Pro
         35                  40                  45
Gly Gln Pro Trp Cys Glu Ala Gln Val Phe Leu Asn Lys Asn Leu
         50                  55                  60
Phe Leu Gln Tyr Asn Ser Asp Asn Asn Met Val Lys Pro Leu Gly
         65                  70                  75
Leu Leu Gly Lys Lys Val Asn Ala Thr Ser Thr Trp Gly Glu Leu
         80                  85                  90
Thr Gln Thr Leu Gly Glu Val Gly Arg Asp Leu Arg Met Leu Leu
         95                  100                 105
Cys Asp Ile Lys Pro Gln Ile Lys Thr Ser Asp Pro Ser Thr Leu
         110                 115                 120
Gln Val Glu Met Phe Cys Gln Arg Glu Ala Glu Arg Cys Thr Gly
         125                 130                 135
Ala Ser Trp Gln Phe Ala Thr Asn Gly Glu Lys Ser Leu Leu Phe
         140                 145                 150
Asp Ala Met Asn Met Thr Trp Thr Val Ile Asn His Glu Ala Ser
         155                 160                 165
Lys Ile Lys Glu Thr Trp Lys Lys Asp Arg Gly Leu Glu Lys Tyr
         170                 175                 180
Phe Arg Lys Leu Ser Lys Gly Asp Cys Asp His Trp Leu Arg Glu
         185                 190                 195
Phe Leu Gly His Trp Glu Ala Met Pro Glu Pro Thr Val Ser Pro
         200                 205                 210
Val Asn Ala Ser Asp Ile His Trp Ser Ser Ser Ser Leu Pro Asp
         215                 220                 225
Arg Trp Ile Ile Leu Gly Ala Phe Ile Leu Leu Val Leu Met Gly
         230                 235                 240
Ile Val Leu Ile Cys Val Trp Trp Gln Asn Gly Arg Arg Ser Thr
         245                 250                 255
```

<210> 8
<211> 1278
<212> PRT
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 72830854CD1

<400> 8
Met Pro Pro Pro Gln Tyr Phe Ser Pro Ala Val Ile Leu Pro Ser
1   5   10   15
Leu Ala Ala Pro Leu Pro Pro Ala Ser Pro Ala Leu Pro Leu Gln
                20   25   30
Ala Val Lys Leu Pro His Pro Pro Gly Ala Pro Leu Ala Met Pro
                35   40   45
Cys Arg Thr Ile Val Pro Asn Ala Pro Ala Thr Ile Pro Leu Leu
                50   55   60
Ala Val Ala Pro Pro Gly Val Ala Ala Leu Ser Ile His Ser Ala
                65   70   75
Val Ala Gln Leu Pro Gly Gln Pro Val Tyr Pro Ala Ala Phe Pro
                80   85   90
Gln Met Ala Pro Thr Asp Val Pro Pro Ser Pro His His Thr Val
                95   100   105
Gln Asn Met Arg Ala Thr Pro Pro Gln Pro Ala Leu Pro Pro Gln
                110   115   120
Pro Thr Leu Pro Pro Gln Pro Val Leu Pro Pro Gln Pro Thr Leu
                125   130   135
Pro Pro Gln Pro Val Leu Pro Pro Gln Pro Thr Arg Pro Pro Gln
                140   145   150
Pro Val Leu Pro Pro Gln Pro Met Leu Pro Pro Gln Pro Val Leu
                155   160   165
Pro Pro Gln Pro Ala Leu Pro Val Arg Pro Glu Pro Leu Gln Pro
                170   175   180
His Leu Pro Glu Gln Ala Ala Pro Ala Ala Thr Pro Gly Ser Gln
                185   190   195
Ile Leu Leu Gly His Pro Ala Pro Tyr Ala Val Asp Val Ala Ala
                200   205   210
Gln Val Pro Thr Val Pro Val Pro Pro Ala Ala Val Leu Ser Pro
                215   220   225
Pro Leu Pro Glu Val Leu Leu Pro Ala Ala Pro Glu Leu Leu Pro
                230   235   240
Gln Phe Pro Ser Ser Leu Ala Thr Val Ser Ala Ser Val Gln Ser
                245   250   255
Val Pro Thr Gln Thr Ala Thr Leu Leu Pro Pro Ala Asn Pro Pro
                260   265   270
Leu Pro Gly Gly Pro Gly Ile Ala Ser Pro Cys Pro Thr Val Gln
                275   280   285
Leu Thr Val Glu Pro Val Gln Glu Glu Gln Ala Ser Gln Asp Lys
                290   295   300
Pro Pro Gly Leu Pro Gln Ser Cys Glu Ser Tyr Gly Gly Ser Asp
                305   310   315
Val Thr Ser Gly Lys Glu Leu Ser Asp Ser Cys Glu Gly Ala Phe
                320   325   330
Gly Gly Gly Arg Leu Glu Gly Arg Ala Ala Arg Lys His His Arg
                335   340   345
Arg Ser Thr Arg Ala Arg Ser Arg Gln Glu Arg Ala Ser Arg Pro
                350   355   360
Arg Leu Thr Ile Leu Asn Val Cys Asn Thr Gly Asp Lys Met Val
                365   370   375
Glu Cys Gln Leu Glu Thr His Asn His Lys Met Val Thr Phe Lys
                380   385   390
Phe Asp Leu Asp Gly Asp Ala Pro Asp Glu Ile Ala Thr Tyr Met
                395   400   405
Val Glu His Asp Phe Ile Leu Gln Ala Glu Arg Glu Thr Phe Ile
                410   415   420
Glu Gln Met Lys Asp Val Met Asp Lys Ala Glu Asp Met Leu Ser
                425   430   435
Glu Asp Thr Asp Ala Asp Arg Gly Ser Asp Pro Gly Thr Ser Pro

```
              440                    445                    450
Pro His Leu Ser Thr Cys Gly Leu Gly Thr Gly Glu Glu Ser Arg
              455                    460                    465
Gln Ser Gln Ala Asn Ala Pro Val Tyr Gln Gln Asn Val Leu His
              470                    475                    480
Thr Gly Lys Arg Trp Phe Ile Ile Cys Pro Val Ala Glu His Pro
              485                    490                    495
Ala Pro Glu Ala Pro Glu Ser Ser Pro Pro Leu Pro Leu Ser Ser
              500                    505                    510
Leu Pro Pro Glu Ala Ser Gln Asp Ser Ala Pro Tyr Lys Asp Gln
              515                    520                    525
Leu Ser Ser Lys Glu Gln Pro Ser Phe Leu Ala Ser Gln Gln Leu
              530                    535                    540
Leu Ser Gln Ala Gly Pro Ser Asn Pro Pro Gly Ala Pro Pro Ala
              545                    550                    555
Pro Leu Ala Pro Ser Ser Pro Pro Val Thr Ala Leu Pro Gln Asp
              560                    565                    570
Gly Ala Ala Pro Ala Thr Ser Thr Met Pro Glu Pro Ala Ser Gly
              575                    580                    585
Thr Ala Ser Gln Ala Gly Gly Pro Gly Thr Pro Gln Gly Leu Thr
              590                    595                    600
Ser Glu Leu Glu Thr Ser Gln Pro Leu Ala Glu Thr His Glu Ala
              605                    610                    615
Pro Leu Ala Val Gln Pro Leu Val Val Gly Leu Ala Pro Cys Thr
              620                    625                    630
Pro Ala Pro Glu Ala Ala Ser Thr Arg Asp Ala Ser Ala Pro Arg
              635                    640                    645
Glu Pro Leu Pro Pro Pro Ala Pro Glu Pro Ser Pro His Ser Gly
              650                    655                    660
Thr Pro Gln Pro Ala Leu Gly Gln Pro Ala Pro Leu Leu Pro Ala
              665                    670                    675
Ala Val Gly Ala Val Ser Leu Ala Thr Ser Gln Leu Pro Ser Pro
              680                    685                    690
Pro Leu Gly Pro Thr Val Pro Pro Gln Pro Pro Ser Ala Leu Glu
              695                    700                    705
Ser Asp Gly Glu Gly Pro Pro Pro Arg Val Gly Phe Val Asp Ser
              710                    715                    720
Thr Ile Lys Ser Leu Asp Glu Lys Leu Arg Thr Leu Leu Tyr Gln
              725                    730                    735
Glu His Val Pro Thr Ser Ser Ala Ser Ala Gly Thr Pro Val Glu
              740                    745                    750
Val Gly Asp Arg Asp Phe Thr Leu Glu Pro Leu Arg Gly Asp Gln
              755                    760                    765
Pro Arg Ser Glu Val Cys Gly Gly Asp Leu Ala Leu Pro Pro Val
              770                    775                    780
Pro Lys Glu Ala Val Ser Gly Arg Val Gln Leu Pro Gln Pro Leu
              785                    790                    795
Val Glu Lys Ser Glu Leu Ala Pro Thr Arg Gly Ala Val Met Glu
              800                    805                    810
Gln Gly Thr Ser Ser Ser Met Thr Ala Glu Ser Ser Pro Arg Ser
              815                    820                    825
Met Leu Gly Tyr Asp Arg Asp Gly Arg Gln Val Ala Ser Asp Ser
              830                    835                    840
His Val Val Pro Ser Val Pro Gln Asp Val Pro Ala Phe Val Arg
              845                    850                    855
Pro Ala Arg Val Glu Pro Thr Asp Arg Asp Gly Gly Glu Ala Gly
              860                    865                    870
Glu Ser Ser Ala Glu Pro Pro Pro Ser Asp Met Gly Thr Val Gly
              875                    880                    885
Gly Gln Ala Ser His Pro Gln Thr Leu Gly Ala Arg Ala Leu Gly
              890                    895                    900
Ser Pro Arg Lys Arg Pro Glu Gln Gln Asp Val Ser Ser Pro Ala
              905                    910                    915
```

```
Lys Thr Val Gly Arg Phe Ser Val Val Ser Thr Gln Asp Glu Trp
            920                     925                     930
Thr Leu Ala Ser Pro His Ser Leu Arg Tyr Ser Ala Pro Pro Asp
            935                     940                     945
Val Tyr Leu Asp Glu Ala Pro Ser Ser Pro Asp Val Lys Leu Ala
            950                     955                     960
Val Arg Arg Ala Gln Thr Ala Ser Ser Ile Glu Val Gly Val Gly
            965                     970                     975
Glu Pro Val Ser Ser Asp Ser Gly Asp Glu Gly Pro Arg Ala Arg
            980                     985                     990
Pro Pro Val Gln Lys Gln Ala Ser Leu Pro Val Ser Gly Ser Val
            995                    1000                    1005
Ala Gly Asp Phe Val Lys Lys Ala Thr Ala Phe Leu Gln Arg Pro
           1010                    1015                    1020
Ser Arg Ala Gly Ser Leu Gly Pro Glu Thr Pro Ser Arg Val Gly
           1025                    1030                    1035
Met Lys Val Pro Thr Ile Ser Val Thr Ser Phe His Ser Gln Ser
           1040                    1045                    1050
Ser Tyr Ile Ser Ser Asp Asn Asp Ser Glu Leu Glu Asp Ala Asp
           1055                    1060                    1065
Ile Lys Lys Glu Leu Gln Ser Leu Arg Glu Lys His Leu Lys Glu
           1070                    1075                    1080
Ile Ser Glu Leu Gln Ser Gln Gln Lys Gln Glu Ile Glu Ala Leu
           1085                    1090                    1095
Tyr Arg Arg Leu Gly Lys Pro Leu Pro Pro Asn Val Gly Phe Phe
           1100                    1105                    1110
His Thr Ala Pro Pro Thr Gly Arg Arg Arg Lys Thr Ser Lys Ser
           1115                    1120                    1125
Lys Leu Lys Ala Gly Lys Leu Leu Asn Pro Leu Val Arg Gln Leu
           1130                    1135                    1140
Lys Val Val Ala Ser Ser Thr Gly Ser Ser Thr Ser Ser Leu Ala
           1145                    1150                    1155
Pro Gly Pro Glu Pro Gly Pro Gln Pro Ala Leu His Val Gln Ala
           1160                    1165                    1170
Gln Val Asn Asn Ser Asn Asn Lys Lys Gly Thr Phe Thr Asp Asp
           1175                    1180                    1185
Leu His Lys Leu Val Asp Glu Trp Thr Ser Lys Thr Val Gly Ala
           1190                    1195                    1200
Ala Gln Leu Lys Pro Thr Leu Asn Gln Leu Lys Gln Thr Gln Lys
           1205                    1210                    1215
Leu Gln Asp Met Glu Ala Gln Ala Gly Trp Ala Ala Pro Gly Glu
           1220                    1225                    1230
Ala Arg Ala Met Thr Ala Pro Arg Ala Gly Val Gly Met Pro Arg
           1235                    1240                    1245
Leu Pro Pro Ala Pro Gly Pro Leu Ser Thr Thr Val Ile Pro Gly
           1250                    1255                    1260
Ala Ala Pro Thr Leu Ser Val Pro Thr Pro Asp Pro Glu Ser Glu
           1265                    1270                    1275
Lys Pro Asp
```

```
<210> 9
<211> 541
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3632052CD1

<400> 9
Met Ser Ser Lys Lys Asn Arg Lys Arg Leu Asn Gln Ser Ala Glu
 1               5                  10                      15
```

```
Asn Gly Ser Ser Leu Pro Ser Ala Ala Ser Ser Cys Ala Glu Ala
            20              25              30
Arg Ala Pro Ser Ala Gly Ser Asp Phe Ala Ala Thr Ser Gly Thr
            35              40              45
Leu Thr Val Thr Asn Leu Leu Glu Lys Gly Lys Glu Phe Arg Val
            50              55              60
Tyr Thr Ala Trp Pro Met Ala Gly Phe Pro Gly Gly Lys Val Gly
            65              70              75
Leu Ser Glu Met Ala Gln Lys Asn Val Gly Val Arg Pro Gly Asp
            80              85              90
Ala Ile Gln Val Gln Pro Leu Val Gly Ala Val Leu Gln Ala Glu
            95              100             105
Glu Met Asp Val Ala Leu Ser Asp Lys Asp Met Glu Ile Asn Glu
            110             115             120
Glu Glu Leu Thr Gly Cys Ile Leu Arg Lys Leu Asp Gly Lys Ile
            125             130             135
Val Leu Pro Gly Asn Phe Leu Tyr Cys Thr Phe Tyr Gly Arg Pro
            140             145             150
Tyr Lys Leu Gln Val Leu Arg Val Lys Gly Ala Asp Gly Met Ile
            155             160             165
Leu Gly Gly Pro Gln Ser Asp Ser Asp Thr Asp Ala Gln Arg Met
            170             175             180
Ala Phe Glu Gln Ser Ser Met Glu Thr Ser Ser Leu Glu Leu Ser
            185             190             195
Leu Gln Leu Ser Gln Leu Asp Leu Glu Asp Thr Gln Ile Pro Thr
            200             205             210
Ser Arg Ser Thr Pro Tyr Lys Pro Ile Asp Asp Arg Ile Thr Asn
            215             220             225
Lys Ala Ser Asp Val Leu Leu Asp Val Thr Gln Ser Pro Gly Asp
            230             235             240
Gly Ser Gly Leu Met Leu Glu Glu Val Thr Gly Leu Lys Cys Asn
            245             250             255
Phe Glu Ser Ala Arg Glu Gly Asn Glu Gln Leu Thr Glu Glu Glu
            260             265             270
Arg Leu Leu Lys Phe Ser Ile Gly Ala Lys Cys Asn Thr Asp Thr
            275             280             285
Phe Tyr Phe Ile Ser Ser Thr Thr Arg Val Asn Phe Thr Glu Ile
            290             295             300
Asp Lys Asn Ser Lys Glu Gln Asp Asn Gln Phe Lys Val Thr Tyr
            305             310             315
Asp Met Ile Gly Gly Leu Ser Ser Gln Leu Lys Ala Ile Arg Glu
            320             325             330
Ile Ile Glu Leu Pro Leu Lys Gln Pro Glu Leu Phe Lys Ser Tyr
            335             340             345
Gly Ile Pro Ala Pro Arg Gly Leu Leu Leu Tyr Gly Pro Pro Cys
            350             355             360
Thr Gly Lys Thr Met Ile Ala Arg Ala Val Ala Asn Glu Val Gly
            365             370             375
Ala Tyr Val Ser Val Ile Asn Gly Pro Glu Ile Ile Ser Lys Phe
            380             385             390
Tyr Gly Glu Thr Glu Ala Lys Leu Arg Gln Ile Phe Ala Glu Ala
            395             400             405
Thr Leu Arg His Pro Ser Ile Ile Phe Ile Asp Glu Leu Asp Ala
            410             415             420
Leu Cys Pro Lys Arg Glu Gly Ala Gln Asn Glu Val Glu Lys Arg
            425             430             435
Val Val Ala Ser Leu Leu Thr Leu Met Asp Gly Ile Gly Ser Glu
            440             445             450
Val Ser Glu Gly Gln Val Leu Val Leu Gly Ala Thr Asn Arg Pro
            455             460             465
His Ala Leu Asp Ala Ala Leu Arg Arg Pro Gly Arg Phe Asp Lys
            470             475             480
Glu Ile Glu Ile Gly Val Pro Asn Ala Gln Asp Arg Leu Asp Ile
```

```
                    485                   490                   495
Leu Gln Lys Leu Leu Arg Arg Val Pro His Leu Leu Thr Glu Ala
                500                   505                   510
Glu Leu Leu Gln Leu Ala Asn Ser Ala His Gly Tyr Val Gly Ala
                515                   520                   525
Asp Leu Lys Val Leu Cys Asn Glu Ala Gly Glu Cys Gly Leu Leu
                530                   535                   540
Trp
```

<210> 10
<211> 553
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7493817CD1

<400> 10
```
Met Pro Ile Leu Leu Phe Leu Ile Asp Thr Ser Ala Ser Met Asn
 1               5                  10                   15
Gln Arg Thr Asp Leu Gly Thr Ser Tyr Leu Asp Ile Ala Lys Gly
                20                  25                   30
Ala Val Glu Leu Phe Leu Lys Leu Arg Ala Arg Asp Pro Ala Ser
                35                  40                   45
Arg Gly Asp Arg Tyr Met Leu Val Thr Tyr Asp Glu Pro Pro Tyr
                50                  55                   60
Cys Ile Lys Ala Gly Trp Lys Glu Asn His Ala Thr Phe Met Ser
                65                  70                   75
Glu Leu Lys Asn Leu Gln Ala Ser Gly Leu Thr Thr Leu Gly Gln
                80                  85                   90
Ala Leu Arg Ser Ser Phe Asp Leu Leu Asn Leu Asn Arg Leu Ile
                95                  100                  105
Ser Gly Ile Asp Asn Tyr Gly Gln Gly Arg Asn Pro Phe Phe Leu
                110                 115                  120
Glu Pro Ser Ile Leu Ile Thr Ile Thr Asp Gly Asn Lys Leu Thr
                125                 130                  135
Ser Thr Ala Gly Val Gln Glu Glu Leu His Pro Pro Leu Asn Ser
                140                 145                  150
Pro Leu Pro Gly Arg Glu Leu Thr Lys Glu Pro Phe Arg Trp Asp
                155                 160                  165
Gln Arg Leu Phe Ala Leu Val Leu Arg Leu Pro Gly Val Ala Ser
                170                 175                  180
Thr Glu Pro Glu Gln Leu Gly Ser Val Pro Thr Asp Glu Ser Ala
                185                 190                  195
Ile Thr Gln Met Cys Glu Val Thr Gly Gly Arg Ser Tyr Cys Val
                200                 205                  210
Arg Thr Gln Arg Met Leu Asn Gln Cys Leu Glu Ser Leu Val Gln
                215                 220                  225
Lys Val Gln Ser Gly Val Val Ile Asn Phe Glu Lys Thr Gly Pro
                230                 235                  240
Asp Pro Leu Pro Ile Gly Glu Asp Gly Leu Met Asp Ser Ser Arg
                245                 250                  255
Pro Ser Asn Ser Phe Ala Ala Gln Pro Trp His Ser Cys His Lys
                260                 265                  270
Leu Ile Tyr Val Arg Pro Asn Ser Lys Thr Gly Val Pro Val Gly
                275                 280                  285
His Trp Pro Ile Pro Glu Ser Phe Trp Pro Asp Gln Asn Leu Pro
                290                 295                  300
Ser Leu Pro Pro Arg Thr Ser His Pro Val Val Arg Phe Ser Cys
                305                 310                  315
Val Asp Cys Glu Pro Met Val Ile Asp Lys Leu Pro Phe Asp Lys
```

```
              320                    325                    330
Tyr Glu Leu Glu Pro Ser Pro Leu Thr Gln Tyr Ile Leu Glu Arg
              335                    340                    345
Lys Ser Pro His Thr Cys Trp Gln Val Phe Val Thr Ser Ser Gly
              350                    355                    360
Lys Tyr Asn Glu Leu Gly Tyr Pro Phe Gly Tyr Leu Lys Ala Ser
              365                    370                    375
Thr Thr Leu Thr Cys Val Asn Leu Phe Val Met Pro Tyr Asn Tyr
              380                    385                    390
Pro Val Leu Leu Pro Leu Leu Asp Asp Leu Phe Lys Val His Lys
              395                    400                    405
Leu Lys Pro Asn Leu Lys Trp Arg Gln Ala Phe Asp Ser Tyr Leu
              410                    415                    420
Lys Thr Leu Pro Pro Tyr Tyr Leu Leu Pro Leu Lys Lys Ala Leu
              425                    430                    435
Arg Met Met Gly Ala Pro Asn Leu Ile Ser Asp Asn Leu Asp Cys
              440                    445                    450
Gly Leu Ser Tyr Ser Val Ile Ser Tyr Leu Lys Lys Leu Ser Gln
              455                    460                    465
Gln Thr Lys Leu Glu Ser Glu Arg Ile Leu Ala Ser Val Gly Lys
              470                    475                    480
Lys Pro Pro Gln Glu Ile Gly Ile Lys Val Lys Asn His Ser Gly
              485                    490                    495
Gly Gly Met Ser Leu Thr His Asn Lys Asn Phe Arg Lys Leu Leu
              500                    505                    510
Lys Glu Ile Thr Gly Glu Thr Ala Leu Arg Leu Thr Glu Leu Asn
              515                    520                    525
Thr Lys Glu Phe Ala Gly Phe Gln Ile Gly Leu Leu Asn Lys Val
              530                    535                    540
Asn Cys Asp Ile Asn Ser Leu Tyr Leu Phe Glu Leu Val
              545                    550
```

```
<210> 11
<211> 804
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6715627CD1

<400> 11
Met Glu Glu Pro Gly Ala Thr Pro Gln Pro Tyr Leu Gly Leu Leu
1               5                   10                     15
Leu Glu Glu Leu Arg Arg Val Val Ala Ala Leu Pro Glu Gly Met
                20                  25                     30
Arg Pro Asp Ser Asn Leu Tyr Gly Phe Pro Trp Glu Leu Val Ile
                35                  40                     45
Cys Ala Ala Val Val Gly Phe Phe Ala Val Leu Phe Phe Leu Trp
                50                  55                     60
Arg Ser Phe Arg Ser Val Arg Ser Arg Leu Tyr Val Gly Arg Glu
                65                  70                     75
Lys Lys Leu Ala Leu Met Leu Ser Gly Leu Ile Glu Glu Lys Ser
                80                  85                     90
Lys Leu Leu Glu Lys Phe Ser Leu Val Gln Lys Glu Tyr Glu Gly
                95                  100                    105
Tyr Glu Val Glu Ser Ser Leu Lys Asp Ala Ser Phe Glu Lys Glu
                110                 115                    120
Ala Thr Glu Ala Gln Ser Leu Glu Ala Thr Cys Glu Lys Leu Asn
                125                 130                    135
Arg Ser Asn Ser Glu Leu Glu Asp Glu Ile Leu Cys Leu Glu Lys
                140                 145                    150
Glu Leu Lys Glu Glu Lys Ser Lys His Ser Glu Gln Asp Glu Leu
```

```
                      155                   160                   165
Met Ala Asp Ile Ser Lys Arg Ile Gln Ser Leu Glu Asp Glu Ser
                      170                   175                   180
Lys Ser Leu Lys Ser Gln Val Ala Glu Ala Lys Met Thr Phe Lys
                      185                   190                   195
Ile Phe Gln Met Asn Glu Glu Arg Leu Lys Ile Ala Ile Lys Asp
                      200                   205                   210
Ala Leu Asn Glu Asn Ser Gln Leu Gln Glu Ser Gln Lys Gln Pro
                      215                   220                   225
Leu Gln Glu Ala Glu Val Trp Lys Glu Gln Val Ser Glu Leu Asn
                      230                   235                   240
Lys Gln Lys Val Thr Phe Glu Asp Ser Lys Val His Ala Glu Gln
                      245                   250                   255
Val Leu Asn Asp Lys Glu Ser His Ile Lys Thr Leu Thr Glu Arg
                      260                   265                   270
Leu Leu Lys Met Lys Asp Trp Ala Ala Met Leu Gly Glu Asp Ile
                      275                   280                   285
Thr Asp Asp Asp Asn Leu Glu Leu Glu Met Asn Ser Glu Ser Glu
                      290                   295                   300
Asn Gly Ala Tyr Leu Asp Asn Pro Pro Lys Gly Ala Leu Lys Lys
                      305                   310                   315
Leu Ile His Ala Ala Lys Leu Asn Ala Ser Leu Lys Thr Leu Glu
                      320                   325                   330
Gly Glu Arg Asn Gln Ile Tyr Ile Gln Leu Ser Glu Val Asp Lys
                      335                   340                   345
Thr Lys Glu Glu Leu Thr Glu His Ile Lys Asn Leu Gln Thr Glu
                      350                   355                   360
Gln Ala Ser Leu Gln Ser Glu Asn Thr His Phe Glu Asn Glu Asn
                      365                   370                   375
Gln Lys Leu Gln Gln Lys Leu Lys Val Met Thr Glu Leu Tyr Gln
                      380                   385                   390
Glu Asn Glu Met Lys Leu His Arg Lys Leu Thr Val Glu Glu Asn
                      395                   400                   405
Tyr Arg Leu Glu Lys Glu Glu Lys Leu Ser Lys Val Asp Glu Lys
                      410                   415                   420
Ile Ser His Ala Thr Glu Glu Leu Glu Thr Tyr Arg Lys Arg Ala
                      425                   430                   435
Lys Asp Leu Glu Glu Glu Leu Glu Arg Thr Ile His Ser Tyr Gln
                      440                   445                   450
Gly Gln Ile Ile Ser His Glu Lys Lys Ala His Asp Asn Trp Leu
                      455                   460                   465
Ala Ala Arg Asn Ala Glu Arg Asn Leu Asn Asp Leu Arg Lys Glu
                      470                   475                   480
Asn Ala His Asn Arg Gln Lys Leu Thr Glu Thr Glu Leu Lys Phe
                      485                   490                   495
Glu Leu Leu Glu Lys Asp Pro Tyr Ala Leu Asp Val Pro Asn Thr
                      500                   505                   510
Ala Phe Gly Arg Glu His Ser Pro Tyr Gly Pro Ser Pro Leu Gly
                      515                   520                   525
Trp Pro Ser Ser Glu Thr Arg Ala Phe Leu Ser Pro Pro Thr Leu
                      530                   535                   540
Leu Glu Gly Pro Leu Arg Leu Ser Pro Leu Leu Pro Gly Gly Gly
                      545                   550                   555
Gly Arg Gly Ser Arg Gly Pro Gly Asn Pro Leu Asp His Gln Ile
                      560                   565                   570
Thr Asn Glu Arg Gly Glu Ser Ser Cys Asp Arg Leu Thr Asp Pro
                      575                   580                   585
His Arg Ala Pro Ser Asp Thr Gly Ser Leu Ser Pro Pro Trp Asp
                      590                   595                   600
Gln Asp Arg Arg Met Met Phe Pro Pro Gly Gln Ser Tyr Pro
                      605                   610                   615
Asp Ser Ala Leu Pro Pro Gln Arg Gln Asp Arg Phe Cys Ser Asn
                      620                   625                   630
```

```
Ser Gly Arg Leu Ser Gly Pro Ala Glu Leu Arg Ser Phe Asn Met
            635               640               645
Pro Ser Leu Asp Lys Met Asp Gly Ser Met Pro Ser Glu Met Glu
            650              .655               660
Ser Ser Arg Asn Asp Thr Lys Asp Asp Leu Gly Asn Leu Asn Val
            665               670               675
Pro Asp Ser Ser Leu Pro Ala Glu Asn Glu Ala Thr Gly Pro Gly
            680               685               690
Leu Phe Leu His Leu Leu Leu Pro Ile Arg Gly Pro Leu Phe Pro
            695               700               705
Val Asp Ala Arg Gly Pro Phe Leu Arg Arg Gly Pro Pro Phe Pro
            710        .         715               720
Pro Pro Pro Pro Gly Ala Met Phe Gly Ala Ser Arg Asp Tyr Phe
            725               730               735
Pro Pro Gly Asp Phe Pro Gly Pro Pro Pro Ala Pro Phe Ala Met
            740               745               750
Arg Asn Val Tyr Pro Pro Arg Gly Phe Pro Pro Tyr Leu Pro Pro
            755               760               765
Arg Pro Gly Phe Phe Pro Pro Pro His Ser Glu Gly Arg Ser
            770               775               780
Glu Phe Pro Ser Gly Leu Ile Pro Pro Ser Asn Glu Pro Ala Thr
            785               790               795
Glu His Pro Glu Pro Gln Gln Glu Thr
            800
```

<210> 12
<211> 465
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7727886CD1

<400> 12

```
Met Arg Arg Ala Ala Gly Met Glu Asp Phe Ser Ala Glu Glu Glu
 1               5                10                15
Glu Ser Trp Tyr Asp Gln Gln Asp Leu Glu Gln Asp Leu His Leu
            20                25                30
Ala Ala Glu Leu Gly Lys Thr Leu Leu Glu Arg Asn Lys Glu Leu
            35                40                45
Glu Gly Ser Leu Gln Gln Met Tyr Ser Thr Asn Glu Glu Gln Val
            50                55                60
Gln Glu Ile Glu Tyr Leu Thr Lys Gln Leu Asp Thr Leu Arg His
            65                70                75
Val Asn Glu Gln His Ala Lys Val Tyr Glu Gln Leu Asp Leu Thr
            80                85                90
Ala Arg Asp Leu Glu Leu Thr Asn His Arg Leu Val Leu Glu Ser
            95               100               105
Lys Ala Ala Gln Gln Lys Ile His Gly Leu Thr Glu Thr Ile Glu
            110               115               120
Arg Leu Gln Ala Gln Val Glu Glu Leu Gln Ala Gln Val Glu Gln
            125               130               135
Leu Arg Gly Leu Glu Gln Leu Arg Val Leu Arg Glu Lys Arg Glu
            140               145               150
Arg Arg Arg Thr Ile His Thr Phe Pro Cys Leu Lys Glu Leu Cys
            155               160               165
Thr Ser Pro Arg Cys Lys Asp Ala Phe Arg Leu His Ser Ser Ser
            170               175               180
Leu Glu Leu Gly Pro Arg Pro Leu Glu Gln Glu Asn Glu Arg Leu
            185               190               195
Gln Thr Leu Val Gly Ala Leu Arg Ser Gln Val Ser Gln Glu Arg
            200               205               210
```

```
Gln Arg Lys Glu Arg Ala Glu Arg Glu Tyr Thr Ala Val Leu Gln
            215                 220                 225
Glu Tyr Ser Glu Leu Glu Arg Gln Leu Cys Glu Met Glu Ala Cys
            230                 235                 240
Arg Leu Arg Val Gln Glu Leu Glu Ala Glu Leu Leu Glu Leu Gln
            245                 250                 255
Gln Met Lys Gln Ala Lys Thr Tyr Leu Leu Gly Pro Asp Asp His
            260                 265                 270
Leu Ala Glu Ala Leu Leu Ala Pro Leu Thr Gln Ala Pro Glu Ala
            275                 280                 285
Asp Asp Pro Gln Pro Gly Arg Gly Asp Asp Leu Gly Ala Gln Asp
            290                 295                 300
Gly Val Ser Ser Pro Ala Ala Ser Pro Gly His Val Val Arg Lys
            305                 310                 315
Ser Cys Ser Asp Thr Ala Leu Asn Ala Ile Val Ala Lys Asp Pro
            320                 325                 330
Ala Ser Arg His Ala Gly Asn Leu Thr Leu His Ala Asn Ser Val
            335                 340                 345
Arg Lys Arg Gly Met Ser Ile Leu Arg Glu Val Asp Glu Gln Tyr
            350                 355                 360
His Ala Leu Leu Glu Lys Tyr Glu Glu Leu Leu Ser Lys Cys Arg
            365                 370                 375
Gln His Gly Ala Gly Val Arg His Ala Gly Val Gln Thr Ser Arg
            380                 385                 390
Pro Ile Ser Arg Asp Ser Ser Trp Arg Asp Leu Arg Gly Gly Glu
            395                 400                 405
Glu Gly Gln Gly Glu Val Lys Ala Gly Glu Lys Ser Leu Ser Gln
            410                 415                 420
His Val Glu Ala Val Asp Lys Arg Leu Glu Gln Ser Gln Pro Glu
            425                 430                 435
Tyr Lys Ala Leu Phe Lys Glu Ile Phe Ser Arg Ile Gln Lys Thr
            440                 445                 450
Lys Ala Asp Ile Asn Ala Thr Lys Val Lys Thr His Ser Ser Lys
            455                 460                 465


<210> 13
<211> 651
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 914113CD1

<400> 13
Met Ser Thr Ala Pro Ser Leu Ser Ala Leu Arg Ser Ser Lys His
  1           5                  10                  15
Ser Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
            20                  25                  30
Gly Ala Asp Pro Ala Trp Thr Ser Ala Leu Ser Gly Asn Ser Ser
            35                  40                  45
Gly Pro Gly Pro Gly Ser Ser Pro Ala Gly Ser Thr Lys Pro Phe
            50                  55                  60
Val His Ala Val Pro Pro Ser Asp Pro Leu Arg Gln Ala Asn Arg
            65                  70                  75
Leu Pro Ile Lys Val Leu Lys Met Leu Thr Ala Arg Thr Gly His
            80                  85                  90
Ile Leu His Pro Glu Tyr Leu Gln Pro Leu Pro Ser Thr Pro Val
            95                  100                 105
Ser Pro Ile Glu Leu Asp Ala Lys Lys Ser Pro Leu Ala Leu Leu
            110                 115                 120
Ala Gln Thr Cys Ser Gln Ile Gly Lys Pro Asp Pro Ser Pro Ser
```

```
              125                    130                    135
Ser Lys Leu Ser Ser Val Ala Ser Asn Gly Gly Gly Ala Gly Gly
              140                    145                    150
Ala Gly Gly Gly Ala Ala Gly Asp Lys Asp Thr Lys Ser Gly Pro
              155                    160                    165
Leu Lys Leu Ser Asp Ile Gly Val Glu Asp Lys Ser Ser Phe Lys
              170                    175                    180
Pro Tyr Ser Lys Pro Gly Ser Asp Lys Lys Glu Pro Gly Gly Gly
              185                    190                    195
Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Val Ser
              200                    205                    210
Ser Glu Lys Ser Gly Phe Arg Val Pro Ser Ala Thr Cys Gln Pro
              215                    220                    225
Phe Thr Pro Arg Thr Gly Ser Pro Glu Leu Gln Arg Leu Gly Leu
              230                    235                    240
Leu Ala Gly Arg Tyr Ala Val Leu Gly Arg Gly Cys Pro Gly Gly
              245                    250                    255
Gln Gly Arg Gln Glu Arg His Arg Arg Gly Arg Pro Val Ala Arg
              260                    265                    270
Ala Pro Gly Gly Ala Ser Ala Glu Gly Gly Pro Thr Gly Leu Ala
              275                    280                    285
His Gly Arg Ile Ser Cys Gly Gly Gly Ile Asn Val Asp Val Asn
              290                    295                    300
Gln His Pro Asp Gly Gly Pro Gly Gly Lys Ala Leu Gly Ser Asp
              305                    310                    315
Cys Gly Gly Ser Ser Gly Ser Ser Ser Gly Ser Gly Pro Ser Ala
              320                    325                    330
Pro Thr Ser Ser Ser Val Leu Gly Ser Gly Leu Val Ala Pro Val
              335                    340                    345
Ser Pro Tyr Lys Pro Gly Gln Thr Val Phe Pro Leu Pro Pro Ala
              350                    355                    360
Gly Met Thr Tyr Pro Gly Ser Leu Ala Gly Ala Tyr Ala Gly Tyr
              365                    370                    375
Pro Pro Gln Phe Leu Pro His Gly Val Ala Leu Asp Pro Thr Lys
              380                    385                    390
Pro Gly Ser Leu Val Gly Ala Gln Leu Ala Ala Ala Ala Ala Gly
              395                    400                    405
Ser Leu Gly Cys Ser Lys Pro Ala Gly Ser Ser Pro Leu Ala Gly
              410                    415                    420
Ala Ser Pro Pro Ser Val Met Thr Ala Ser Leu Cys Arg Asp Pro
              425                    430                    435
Tyr Cys Leu Ser Tyr His Cys Ala Ser His Leu Ala Gly Ala Ala
              440                    445                    450
Ala Ala Ser Ala Ser Cys Ala His Asp Pro Ala Ala Ala Ala Ala
              455                    460                    465
Ala Leu Lys Ser Gly Tyr Pro Leu Val Tyr Pro Thr His Pro Leu
              470                    475                    480
His Gly Val His Ser Ser Leu Thr Ala Ala Ala Ala Gly Ala
              485                    490                    495
Thr Pro Pro Ser Leu Ala Gly His Pro Leu Tyr Pro Tyr Gly Phe
              500                    505                    510
Met Leu Pro Asn Asp Pro Leu Pro His Ile Cys Asn Trp Val Ser
              515                    520                    525
Ala Asn Gly Pro Cys Asp Lys Arg Phe Ala Thr Ser Glu Glu Leu
              530                    535                    540
Leu Ser His Leu Arg Thr His Thr Ala Phe Pro Gly Thr Asp Lys
              545                    550                    555
Leu Leu Ser Gly Tyr Pro Ser Ser Ser Leu Ala Ser Ala Ala
              560                    565                    570
Ala Ala Ala Met Ala Cys His Met His Ile Pro Thr Ser Gly Ala
              575                    580                    585
Pro Gly Ser Pro Gly Thr Leu Ala Leu Arg Ser Pro His His Ala
              590                    595                    600
```

Leu Gly Leu Ser Ser Arg Tyr His Pro Tyr Ser Lys Ser Pro Leu
                605   610   615
Pro Thr Pro Gly Ala Pro Val Pro Val Pro Ala Ala Thr Gly Pro
                620   625   630
Tyr Tyr Ser Pro Tyr Ala Leu Tyr Gly Gln Arg Leu Thr Thr Ala
                635   640   645
Ser Ala Leu Gly Tyr Gln
                650

<210> 14
<211> 1443
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1953711CD1

<400> 14
Met Ala Thr Ser Arg Gly Ala Ser Arg Cys Pro Arg Asp Ile Ala
 1   5   10   15
Asn Val Met Gln Arg Leu Gln Asp Glu Gln Glu Ile Val Gln Lys
                20   25   30
Arg Thr Phe Thr Lys Trp Ile Asn Ser His Leu Ala Lys Arg Lys
                35   40   45
Pro Pro Met Val Val Asp Asp Leu Phe Glu Asp Met Lys Asp Gly
                50   55   60
Val Lys Leu Leu Ala Leu Leu Glu Val Leu Ser Gly Gln Lys Leu
                65   70   75
Pro Cys Glu Gln Gly Arg Arg Met Lys Arg Ile His Ala Val Ala
                80   85   90
Asn Ile Gly Thr Ala Leu Lys Phe Leu Glu Gly Arg Lys Ile Lys
                95   100   105
Leu Val Asn Ile Asn Ser Thr Asp Ile Ala Asp Gly Arg Pro Ser
                110   115   120
Ile Val Leu Gly Leu Met Trp Thr Ile Ile Leu Tyr Phe Gln Ile
                125   130   135
Glu Glu Leu Thr Ser Asn Leu Pro Gln Leu Gln Ser Leu Ser Ser
                140   145   150
Ser Ala Ser Ser Val Asp Ser Ile Val Ser Ser Glu Thr Pro Ser
                155   160   165
Pro Pro Ser Lys Arg Lys Val Thr Thr Lys Ile Gln Gly Asn Ala
                170   175   180
Lys Lys Ala Leu Leu Lys Trp Val Gln Tyr Thr Ala Gly Lys Gln
                185   190   195
Thr Gly Ile Glu Val Lys Asp Phe Gly Lys Ser Trp Arg Ser Gly
                200   205   210
Val Ala Phe His Ser Val Ile His Ala Ile Arg Pro Glu Leu Val
                215   220   225
Asp Leu Glu Thr Val Lys Gly Arg Ser Asn Arg Glu Asn Leu Glu
                230   235   240
Asp Ala Phe Thr Ile Ala Glu Thr Glu Leu Gly Ile Pro Arg Leu
                245   250   255
Leu Asp Pro Glu Asp Val Asp Val Asp Lys Pro Asp Glu Lys Ser
                260   265   270
Ile Met Thr Tyr Val Ala Gln Phe Leu Lys His Tyr Pro Asp Ile
                275   280   285
His Asn Ala Ser Thr Asp Gly Gln Glu Asp Asp Glu Ile Leu Pro
                290   295   300
Gly Phe Pro Ser Phe Ala Asn Ser Val Gln Asn Phe Lys Arg Glu
                305   310   315
Asp Arg Val Ile Phe Lys Glu Met Lys Val Trp Ile Glu Gln Phe
                320   325   330

```
Glu Arg Asp Leu Thr Arg Ala Gln Met Val Glu Ser Asn Leu Gln
        335                 340                 345
Asp Lys Tyr Gln Ser Phe Lys His Phe Arg Val Gln Tyr Glu Met
        350                 355                 360
Lys Arg Lys Gln Ile Glu His Leu Ile Gln Pro Leu His Arg Asp
        365                 370                 375
Gly Lys Leu Ser Leu Asp Gln Ala Leu Val Lys Gln Ser Trp Asp
        380                 385                 390
Arg Val Thr Ser Arg Leu Phe Asp Trp His Ile Gln Leu Asp Lys
        395                 400                 405
Ser Leu Pro Ala Pro Leu Gly Thr Ile Gly Ala Trp Leu Tyr Arg
        410                 415                 420
Ala Glu Val Ala Leu Arg Glu Glu Ile Thr Val Gln Gln Val His
        425                 430                 435
Glu Glu Thr Ala Asn Thr Ile Gln Arg Lys Leu Glu Gln His Lys
        440                 445                 450
Asp Leu Leu Gln Asn Thr Asp Ala His Lys Arg Ala Phe His Glu
        455                 460                 465
Ile Tyr Arg Thr Arg Ser Val Asn Gly Ile Pro Val Pro Pro Asp
        470                 475                 480
Gln Leu Glu Asp Met Ala Glu Arg Phe His Phe Val Ser Ser Thr
        485                 490                 495
Ser Glu Leu His Leu Met Lys Met Glu Phe Leu Glu Leu Lys Tyr
        500                 505                 510
Arg Leu Leu Ser Leu Leu Val Leu Ala Glu Ser Lys Leu Lys Ser
        515                 520                 525
Trp Ile Ile Lys Tyr Gly Arg Arg Glu Ser Val Glu Gln Leu Leu
        530                 535                 540
Gln Asn Tyr Val Ser Phe Ile Glu Asn Ser Lys Phe Phe Glu Gln
        545                 550                 555
Tyr Glu Val Thr Tyr Gln Ile Leu Lys Gln Thr Ala Glu Met Tyr
        560                 565                 570
Val Lys Ala Asp Gly Ser Val Glu Glu Ala Glu Asn Val Ile Lys
        575                 580                 585
Phe Met Asn Glu Thr Thr Ala Gln Trp Arg Asn Leu Ser Val Glu
        590                 595                 600
Val Arg Ser Val Arg Ser Met Leu Glu Glu Val Ile Ser Asn Trp
        605                 610                 615
Asp Arg Tyr Gly Asn Thr Val Ala Ser Leu Gln Ala Trp Leu Glu
        620                 625                 630
Asp Ala Glu Lys Met Leu Asn Gln Ser Glu Asn Ala Lys Lys Asp
        635                 640                 645
Phe Phe Arg Asn Leu Pro His Trp Ile Gln His His Thr Ala Met
        650                 655                 660
Asn Asp Ala Gly Asn Phe Leu Ile Glu Thr Cys Asp Glu Met Val
        665                 670                 675
Ser Arg Asp Leu Lys Gln Gln Leu Leu Leu Leu Asn Gly Arg Trp
        680                 685                 690
Arg Glu Leu Phe Met Gly Val Lys Gln Tyr Ala Gln Ala Asp Glu
        695                 700                 705
Met Asp Arg Met Lys Lys Glu Tyr Thr Asp Cys Val Val Thr Leu
        710                 715                 720
Ser Ala Phe Ala Thr Glu Ala His Lys Lys Leu Ser Glu Pro Leu
        725                 730                 735
Glu Val Ser Phe Met Asn Val Lys Leu Leu Ile Gln Asp Leu Glu
        740                 745                 750
Asp Ile Glu Gln Arg Val Pro Val Met Asp Ala Gln Tyr Lys Ile
        755                 760                 765
Ile Thr Lys Thr Ala His Leu Ile Thr Lys Glu Ser Pro Gln Glu
        770                 775                 780
Glu Gly Lys Glu Met Phe Ala Thr Met Ser Lys Leu Lys Glu Gln
        785                 790                 795
Leu Thr Lys Val Arg Glu Cys Tyr Ser Pro Leu Leu Tyr Glu Ser
```

                                23/71

```
                    800                    805                    810
Gln Gln Leu Leu Ile Pro Leu Glu Glu Leu Glu Lys Gln Met Thr
                815                    820                    825
Ser Phe Tyr Asp Ser Leu Gly Lys Ile Asn Glu Ile Ile Thr Val
                830                    835                    840
Leu Glu Arg Glu Ala Gln Ser Ser Ala Leu Phe Lys Gln Lys His
                845                    850                    855
Gln Glu Leu Leu Ala Cys Gln Glu Asn Cys Lys Lys Thr Leu Thr
                860                    865                    870
Leu Ile Glu Lys Gly Ser Gln Ser Val Gln Lys Phe Val Thr Leu
                875                    880                    885
Ser Asn Val Leu Lys His Phe Asp Gln Thr Arg Leu Gln Arg Gln
                890                    895                    900
Ile Ala Asp Ile His Val Ala Phe Gln Ser Met Val Lys Lys Thr
                905                    910                    915
Gly Asp Trp Lys Lys His Val Glu Thr Asn Ser Arg Leu Met Lys
                920                    925                    930
Lys Phe Glu Glu Ser Arg Ala Glu Leu Glu Lys Val Leu Arg Ile
                935                    940                    945
Ala Gln Glu Gly Leu Glu Glu Lys Gly Asp Pro Glu Glu Leu Leu
                950                    955                    960
Arg Arg His Thr Glu Phe Phe Ser Gln Leu Asp Gln Arg Val Leu
                965                    970                    975
Asn Ala Phe Leu Lys Ala Cys Asp Glu Leu Thr Asp Ile Leu Pro
                980                    985                    990
Glu Gln Glu Gln Gln Gly Leu Gln Glu Ala Val Arg Lys Leu His
                995                    1000                   1005
Lys Gln Trp Lys Asp Leu Gln Gly Glu Ala Pro Tyr His Leu Leu
                1010                   1015                   1020
His Leu Lys Ile Asp Val Glu Lys Asn Arg Phe Leu Ala Ser Val
                1025                   1030                   1035
Glu Glu Cys Arg Thr Glu Leu Asp Arg Glu Thr Lys Leu Met Pro
                1040                   1045                   1050
Gln Glu Gly Ser Glu Lys Ile Ile Lys Glu His Arg Val Phe Phe
                1055                   1060                   1065
Ser Asp Lys Gly Pro His His Leu Cys Glu Lys Arg Leu Gln Leu
                1070                   1075                   1080
Ile Glu Glu Leu Cys Val Lys Leu Pro Val Arg Asp Pro Val Arg
                1085                   1090                   1095
Asp Thr Pro Gly Thr Cys His Val Thr Leu Lys Glu Leu Arg Ala
                1100                   1105                   1110
Ala Ile Asp Ser Thr Tyr Arg Lys Leu Met Glu Asp Pro Asp Lys
                1115                   1120                   1125
Trp Lys Asp Tyr Thr Ser Arg Phe Ser Glu Phe Ser Ser Trp Ile
                1130                   1135                   1140
Ser Thr Asn Glu Thr Gln Leu Lys Gly Ile Lys Gly Glu Ala Ile
                1145                   1150                   1155
Asp Thr Ala Asn His Gly Glu Val Lys Arg Ala Val Glu Glu Ile
                1160                   1165                   1170
Arg Asn Gly Val Thr Lys Arg Gly Glu Thr Leu Ser Trp Leu Lys
                1175                   1180                   1185
Ser Arg Leu Lys Val Leu Thr Glu Val Ser Ser Glu Asn Glu Ala
                1190                   1195                   1200
Gln Lys Gln Gly Asp Glu Leu Ala Lys Leu Ser Ser Ser Phe Lys
                1205                   1210                   1215
Ala Leu Val Thr Leu Leu Ser Glu Val Glu Lys Met Leu Ser Asn
                1220                   1225                   1230
Phe Gly Asp Cys Val Gln Tyr Lys Glu Ile Val Lys Asn Ser Leu
                1235                   1240                   1245
Glu Glu Leu Ile Ser Gly Ser Lys Glu Val Gln Glu Gln Ala Glu
                1250                   1255                   1260
Lys Ile Leu Asp Thr Glu Asn Leu Phe Glu Ala Gln Gln Leu Leu
                1265                   1270                   1275
```

```
Leu His His Gln Gln Lys Thr Lys Arg Ile Ser Ala Lys Lys Arg
            1280                1285                1290
Asp Val Gln Gln Gln Ile Ala Gln Ala Gln Gln Gly Glu Gly Gly
            1295                1300                1305
Leu Pro Asp Arg Gly His Glu Glu Leu Arg Lys Leu Glu Ser Thr
            1310                1315                1320
Leu Asp Gly Leu Glu Arg Ser Arg Glu Arg Gln Glu Arg Arg Ile
            1325                1330                1335
Gln Val Thr Leu Arg Lys Trp Glu Arg Phe Glu Thr Asn Lys Glu
            1340                1345                1350
Thr Val Val Arg Tyr Leu Phe Gln Thr Gly Ser Ser His Glu Arg
            1355                1360                1365
Phe Leu Ser Phe Ser Ser Leu Glu Ser Leu Ser Ser Glu Leu Glu
            1370                1375                1380
Gln Thr Lys Glu Phe Ser Lys Arg Thr Glu Ser Ile Ala Val Gln
            1385                1390                1395
Ala Glu Asn Leu Val Lys Glu Ala Ser Glu Ile Pro Leu Gly Pro
            1400                1405                1410
Gln Asn Lys Gln Leu Leu Gln Gln Gln Ala Lys Ser Ile Lys Glu
            1415                1420                1425
Gln Val Lys Lys Leu Glu Asp Thr Leu Glu Glu Glu Tyr Val Ile
            1430                1435                1440
Asp Lys Ser
```

<210> 15
<211> 1087
<212> PRT
<213> Homo sapiens
:

<220>
<221> misc_feature
<223> Incyte ID No: 1595275CD1

<400> 15

```
Met Ala Glu Phe Thr Ser Tyr Lys Glu Thr Ala Ser Ser Arg His
1                 5                 10                15
Leu Arg Phe Lys Leu Gln Ser Leu Ser Arg Arg Leu Asp Glu Leu
            20                25                30
Glu Glu Ala Thr Lys Asn Leu Gln Lys Ala Glu Asp Glu Leu Leu
            35                40                45
Asp Leu Gln Asp Lys Val Ile Gln Ala Glu Gly Ser Asn Ser Ser
            50                55                60
Met Leu Ala Glu Ile Glu Val Leu Arg Gln Arg Val Leu Arg Ile
            65                70                75
Glu Gly Lys Asp Glu Glu Ile Lys Arg Ala Glu Asp Leu Cys Arg
            80                85                90
Leu Met Lys Glu Lys Leu Glu Glu Glu Glu Asn Leu Thr Arg Glu
            95                100               105
Leu Lys Ser Glu Ile Glu Arg Leu Gln Lys Arg Met Ala Glu Leu
            110               115               120
Glu Lys Leu Glu Glu Ala Phe Ser Arg Ser Lys Asn Asp Cys Thr
            125               130               135
Gln Leu Cys Leu Ser Leu Asn Glu Glu Arg Asn Leu Thr Lys Lys
            140               145               150
Ile Ser Ser Glu Leu Glu Met Leu Arg Val Lys Val Lys Glu Leu
            155               160               165
Glu Ser Ser Glu Asp Arg Leu Asp Lys Thr Glu Gln Ser Leu Ala
            170               175               180
Ser Glu Leu Glu Lys Leu Lys Ser Leu Thr Leu Ser Phe Val Ser
            185               190               195
Glu Arg Lys Tyr Leu Asn Glu Lys Glu Lys Glu Asn Glu Lys Leu
            200               205               210
```

```
Ile Lys Glu Leu Thr Gln Lys Leu Glu Gln Asn Lys Lys Met Asn
              215                   220                   225
Arg Asp Tyr Thr Arg Asn Ala Ser Asn Leu Glu Arg Asn Asp Leu
              230                   235                   240
Arg Ile Glu Asp Gly Ile Ser Ser Thr Leu Pro Ser Lys Glu Ser
              245                   250                   255
Arg Arg Lys Gly Gly Leu Asp Tyr Leu Lys Gln Val Glu Asn Glu
              260   ,                 265                   270
Thr Arg Asn Lys Ser Glu Asn Glu Lys Asn Arg Asn Gln Glu Asp
              275                   280                   285
Asn Lys Val Lys Asp Leu Asn Gln Glu Ile Glu Lys Leu Lys Thr
              290                   295                   300
Gln Ile Lys His Phe Glu Ser Leu Glu Glu Glu Leu Lys Lys Met
              305                   310                   315
Lys Ser Lys Asn Asn Asp Leu Gln Asp Asn Tyr Leu Ser Glu Gln
              320                   325                   330
Asn Lys Asn Lys Leu Leu Ala Ser Gln Leu Glu Glu Ile Lys Leu
              335                   340                   345
Gln Ile Lys Lys Gln Lys Glu Leu Glu Asn Gly Glu Val Glu Gly
              350                   355                   360
Glu Asp Ala Phe Leu Ser Ser Lys Gly Arg His Glu Arg Thr Lys
              365                   370                   375
Phe Arg Gly His Gly Ser Glu Ala Ser Val Ser Lys His Thr Ala
              380                   385                   390
Arg Glu Leu Ser Pro Gln His Lys Arg Glu Arg Leu Arg Asn Arg
              395                   400                   405
Glu Phe Ala Leu Asn Asn Glu Asn Tyr Ser Leu Ser Asn Arg Gln
              410                   415                   420
Val Ser Ser Pro Ser Phe Thr Asn Arg Arg Ala Ala Lys Ala Ser
              425                   430                   435
His Met Gly Val Ser Thr Asp Ser Gly Thr Gln Glu Thr Lys Lys
              440                   445                   450
Thr Glu Asp Arg Phe Val Pro Ser Ser Ser Lys Ser Glu Gly Lys
              455                   460                   465
Lys Ser Arg Glu Gln Pro Ser Val Leu Ser Arg Tyr Pro Pro Ala
              470                   475                   480
Ala Gln Glu His Ser Lys Ala Trp Lys Gly Thr Ser Lys Pro Gly
              485                   490                   495
Thr Glu Ser Gly Leu Lys Gly Lys Val Glu Lys Thr Thr Arg Thr
              500                   505                   510
Phe Ser Asp Thr Thr His Gly Ser Val Pro Ser Asp Pro Leu Gly
              515                   520                   525
Arg Ala Asp Lys Ala Ser Asp Thr Ser Ser Glu Thr Val Phe Gly
              530                   535                   540
Lys Arg Gly His Val Leu Gly Asn Gly Ser Gln Val Thr Gln Ala
              545                   550                   555
Ala Asn Ser Gly Cys Ser Lys Ala Ile Gly Ala Leu Ala Ser Ser
              560                   565                   570
Arg Arg Ser Ser Ser Glu Gly Leu Ser Lys Gly Lys Lys Ala Ala
              575                   580                   585
Asn Gly Leu Glu Ala Asp Asn Ser Cys Pro Asn Ser Lys Ala Pro
              590                   595                   600
Val Leu Ser Lys Tyr Pro Tyr Ser Cys Arg Ser Gln Glu Asn Ile
              605                   610                   615
Leu Gln Gly Phe Ser Thr Ser His Lys Glu Gly Val Asn Gln Pro
              620                   625                   630
Ala Ala Val Val Met Glu Asp Ser Ser Pro His Glu Ala Leu Arg
              635                   640                   645
Cys Arg Val Ile Lys Ser Ser Gly Arg Glu Lys Pro Asp Ser Asp
              650                   655                   660
Asp Asp Leu Asp Ile Ala Ser Leu Val Thr Ala Lys Leu Val Asn
             .665                   670                   675
Thr Thr Ile Thr Pro Glu Pro Glu Pro Lys Pro Gln Pro Asn Ser
```

26/71

```
                 680                   685                   690
Arg Glu Lys Ala Lys Thr Arg Gly Ala Pro Arg Thr Ser Leu Phe
                 695                   700                   705
Glu Asn Asp Lys Asp Ala Gly Met Glu Asn Glu Ser Val Lys Ser
                 710                   715                   720
Val Arg Ala Ser Thr Asn Thr Met Glu Leu Pro Asp Thr Asn Gly
                 725                   730                   735
Ala Gly Val Lys Ser Gln Arg Pro Phe Ser Pro Arg Glu Ala Leu
                 740                   745                   750
Arg Ser Arg Ala Ile Ile Lys Pro Val Ile Val Asp Lys Asp Val
                 755                   760                   765
Lys Lys Ile Met Gly Gly Ser Gly Thr Glu Thr Thr Leu Glu Lys
                 770                   775                   780
Gln Lys Pro Val Ser Lys Pro Gly Pro Asn Lys Val Thr Ser Ser
                 785                   790                   795
Ile Thr Ile Tyr Pro Ser Asp Ser Ser Ser Pro Arg Ala Ala Pro
                 800                   805                   810
Gly Glu Ala Leu Arg Glu Arg His Thr Ser Thr Ser Asn Ile Gln
                 815                   820                   825
Val Gly Leu Ala Glu Leu Thr Ser Val Ser Asn His Val Ser Ser
                 830                   835                   840
Pro Phe Glu Leu Ser Ile His Lys His Asp Ile Thr Leu Gln Leu
                 845                   850                   855
Ala Glu Ala Glu Arg Met Ala Asp Gly Pro Leu Lys Asn Arg Pro
                 860                   865                   870
Glu Thr Val Val Ser Arg Ser Ser Ile Ile Ile Lys Pro Ser Asp
                 875                   880                   885
Pro Val Glu Arg Asn Ser His Ala Pro Pro Ala Glu Thr Ile Arg
                 890                   895                   900
Trp Lys Ser His Ser Ala Pro Ser Glu Val Gly Phe Ser Asp Ala
                 905                   910                   915
Arg His Val Thr Val Arg Asn Ala Trp Lys Ser Arg Arg Asp Leu
                 920                   925                   930
Lys Ser Leu Glu Asp Pro Pro Thr Arg Ile Gly Lys Asn Val Glu
                 935                   940                   945
Ser Thr Asn Ser Asn Ala Tyr Thr Gln Arg Ser Ser Thr Asp Phe
                 950                   955                   960
Ser Glu Leu Glu Gln Pro Arg Ser Cys Leu Phe Glu Gln Gly Thr
                 965                   970                   975
Arg Arg Val Gly Pro Ser Ser Gly Asp Ala Pro Glu Pro Ser Ser
                 980                   985                   990
Arg Arg Thr Gln Ser Ser Leu Thr Val Ser Glu Val Leu Thr Arg
                 995                  1000                  1005
Arg Asn Arg Val Gly Asp Thr Ile Thr Val Ala Ala Trp Asn His
                1010                  1015                  1020
Ser Ala Ser Met Glu Glu Glu Gly Glu Asp Cys Thr Leu Ser Val
                1025                  1030                  1035
Tyr Arg Gln Leu His Asn Pro Trp Ile Arg Leu Asn Cys Leu Gly
                1040                  1045                  1050
Ser Arg Gly Cys Gln Ser Leu Gly Glu Tyr Gly Arg Gly Asn Asp
                1055                  1060                  1065
Tyr Gly Pro Asn Arg Pro Cys Ala Glu Glu Thr Leu Ser Pro His
                1070                  1075                  1080
Gly Cys Leu Leu Ser Ala Leu
                1085
```

<210> 16
<211> 240
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 72332548CD1

<400> 16
```
Met Ala Thr Ala Met Tyr Leu Glu His Tyr Leu Asp Ser Ile Glu
1               5                   10                  15
Asn Leu Pro Cys Glu Leu Gln Arg Asn Phe Gln Leu Met Arg Glu
                20                  25                  30
Leu Asp Gln Arg Thr Glu Asp Lys Lys Ala Glu Ile Asp Ile Leu
                35                  40                  45
Ala Ala Glu Tyr Ile Ser Thr Val Lys Thr Leu Ser Pro Asp Gln
                50                  55                  60
Arg Val Glu Arg Leu Gln Lys Ile Gln Asn Ala Tyr Ser Lys Cys
                65                  70                  75
Lys Glu Tyr Ser Asp Asp Lys Val Gln Leu Ala Met Gln Thr Tyr
                80                  85                  90
Glu Met Val Asp Lys His Ile Arg Arg Leu Asp Ala Asp Leu Ala
                95                  100                 105
Arg Phe Glu Ala Asp Leu Lys Asp Lys Met Glu Gly Ser Asp Phe
                110                 115                 120
Glu Ser Ser Gly Gly Arg Gly Leu Lys Lys Gly Arg Gly Gln Lys
                125                 130                 135
Glu Lys Arg Gly Ser Arg Gly Arg Gly Arg Thr Ser Glu Glu
                140                 145                 150
Asp Thr Pro Lys Lys Lys Lys His Lys Gly Gly Ser Glu Phe Thr
                155                 160                 165
Asp Thr Ile Leu Pro Val His Pro Ser Asp Val Leu Asp Met Pro
                170                 175                 180
Val Asp Pro Asn Glu Pro Thr Tyr Cys Leu Cys His Gln Val Ser
                185                 190                 195
Tyr Gly Glu Met Ile Gly Cys Asp Asn Pro Asp Cys Pro Ile Glu
                200                 205                 210
Trp Phe His Phe Ala Cys Val Asp Leu Thr Thr Lys Pro Lys Gly
                215                 220                 225
Lys Trp Phe Cys Pro Arg Cys Val Gln Glu Lys Arg Lys Lys Lys
                230                 235                 240
```

<210> 17
<211> 686
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7322834CD1

<400> 17
```
Met Ala Ala Asn Met Tyr Arg Val Gly Asp Tyr Val Tyr Phe Glu
1               5                   10                  15
Asn Ser Ser Ser Asn Pro Tyr Leu Ile Arg Arg Ile Glu Glu Leu
                20                  25                  30
Asn Lys Thr Ala Asn Gly Asn Val Glu Ala Lys Val Val Cys Phe
                35                  40                  45
Tyr Arg Arg Arg Asp Ile Ser Ser Thr Leu Ile Ala Leu Ala Asp
                50                  55                  60
Lys His Ala Arg Glu Ile Glu Glu Glu Met Glu Asn Pro Glu Met
                65                  70                  75
Val Asp Leu Pro Glu Lys Leu Lys His Gln Leu Arg His Arg Glu
                80                  85                  90
Leu Phe Leu Ser Arg Gln Leu Glu Ser Leu Pro Ala Thr His Ile
                95                  100                 105
Arg Gly Lys Cys Ser Val Thr Leu Leu Asn Glu Thr Glu Ser Leu
                110                 115                 120
```

```
Lys Ser Tyr Leu Glu Arg Glu Asp Phe Phe Phe Tyr Ser Leu Val
            125                 130                 135
Tyr Asp Pro Gln Gln Lys Thr Leu Leu Ala Asp Lys Gly Glu Ile
            140                 145                 150
Arg Val Gly Asn Arg Tyr Gln Ala Asp Ile Thr Asp Leu Leu Lys
            155                 160                 165
Glu Gly Glu Glu Asp Gly Arg Asp Gln Ser Arg Leu Glu Thr Gln
            170                 175                 180
Val Trp Glu Ala His Asn Pro Leu Thr Asp Lys Gln Ile Asp Gln
            185                 190                 195
Phe Leu Val Val Ala Arg Ser Val Gly Thr Phe Ala Arg Ala Leu
            200                 205                 210
Asp Cys Ser Ser Ser Val Arg Gln Pro Ser Leu His Met Ser Ala
            215                 220                 225
Ala Ala Ala Ser Arg Asp Ile Thr Leu Phe His Ala Met Asp Thr
            230                 235                 240
Leu His Lys Asn Ile Tyr Asp Ile Ser Lys Ala Ile Ser Ala Leu
            245                 250                 255
Val Pro Gln Gly Gly Pro Val Leu Cys Arg Asp Glu Met Glu Glu
            260                 265                 270
Trp Ser Ala Ser Glu Ala Asn Leu Phe Glu Glu Ala Leu Glu Lys
            275                 280                 285
Tyr Gly Lys Asp Phe Thr Asp Ile Gln Gln Asp Phe Leu Pro Trp
            290                 295                 300
Lys Ser Leu Thr Ser Ile Ile Glu Tyr Tyr Tyr Met Trp Lys Thr
            305                 310                 315
Thr Asp Arg Tyr Val Gln Gln Lys Arg Leu Lys Ala Ala Glu Ala
            320                 325                 330
Glu Ser Lys Leu Lys Gln Val Tyr Ile Pro Asn Tyr Asn Lys Pro
            335                 340                 345
Asn Pro Asn Gln Ile Ser Val Asn Asn Val Lys Ala Gly Val Val
            350                 355                 360
Asn Gly Thr Gly Ala Pro Gly Gln Ser Pro Gly Ala Gly Arg Ala
            365                 370                 375
Cys Glu Ser Cys Tyr Thr Thr Gln Ser Tyr Gln Trp Tyr Ser Trp
            380                 385                 390
Gly Pro Pro Asn Met Gln Cys Arg Leu Cys Ala Ser Cys Trp Thr
            395                 400                 405
Tyr Trp Lys Lys Tyr Gly Gly Leu Lys Met Pro Thr Arg Leu Asp
            410                 415                 420
Gly Glu Arg Pro Gly Pro Asn Arg Ser Asn Met Ser Pro His Gly
            425                 430                 435
Leu Pro Ala Arg Ser Ser Gly Ser Pro Lys Phe Ala Met Lys Thr
            440                 445                 450
Arg Gln Ala Phe Tyr Leu His Thr Thr Lys Leu Thr Arg Ile Ala
            455                 460                 465
Arg Arg Leu Cys Arg Glu Ile Leu Arg Pro Trp His Ala Ala Arg
            470                 475                 480
His Pro Tyr Leu Pro Ile Asn Ser Ala Ala Ile Lys Ala Glu Cys
            485                 490                 495
Thr Ala Arg Leu Pro Glu Ala Ser Gln Ser Pro Leu Val Leu Lys
            500                 505                 510
Gln Ala Val Arg Lys Pro Leu Glu Ala Val Leu Arg Tyr Leu Glu
            515                 520                 525
Thr His Pro Arg Pro Pro Lys Pro Asp Pro Val Lys Ser Val Ser
            530                 535                 540
Ser Val Leu Ser Ser Leu Thr Pro Ala Lys Val Ala Pro Val Ile
            545                 550                 555
Asn Asn Gly Ser Pro Thr Ile Leu Gly Lys Arg Ser Tyr Glu Gln
            560                 565                 570
His Asn Gly Val Asp Gly Leu Ala Asn His Gly Gln Thr Arg His
            575                 580                 585
Met Gly Pro Ser Arg Asn Leu Leu Leu Asn Gly Lys Ser Tyr Pro
```

```
                590                    595                    600
Thr Lys Val Arg Leu Ile Arg Gly Gly Ser Leu Pro Pro Val Lys
                605                    610                    615
Arg Arg Arg Met Asn Trp Ile Asp Ala Pro Asp Asp Val Phe Tyr
                620                    625                    630
Met Ala Thr Glu Glu Thr Arg Lys Ile Arg Lys Leu Leu Ser Ser
                635                    640                    645
Ser Glu Thr Lys Arg Ala Ala Arg Arg Pro Tyr Lys Pro Ile Ala
                650                    655                    660
Leu Arg Gln Ser Gln Ala Leu Pro Pro Arg Pro Pro Pro Pro·Ala
                665                    670                    675
Pro Val Asn Asp Glu Pro Ile Val Ile Glu Asp
                680                    685
```

```
<210> 18
<211> 1595
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1925714CD1

<400> 18
Met Gly Asp Ile Thr Thr Asn Thr Thr Glu Ile Gln Lys Ile Thr
 1               5                    10                    15
Lys Gly Tyr Tyr Glu His Leu Tyr Val His Lys Leu Glu Tyr Leu
                20                    25                    30
Glu Glu Met Asn Lys Phe Leu Lys Lys Tyr Asn Pro Ser Ser Leu
                35                    40                    45
Asn Gln Lys Glu Leu Glu Ile Leu Asn Arg Pro Ile Thr Ser Ser
                50                    55                    60
Gln Ile Glu Leu Val Ile Lys Lys Leu Pro Thr Gly Trp Thr Pro
                65                    70                    75
Pro Leu Gly Ala Gly His Ser Gln Thr Lys Gly Ser Arg Asn Leu
                80                    85                    90
Cys Arg Leu Lys Cys Pro Cys Leu Thr Ala Val Lys Arg Ile Val
                95                    100                   105
Val Leu Pro Gly Arg Ser Trp Arg Ser Glu Asn Gly Gln Thr Ala
                110                   115                   120
Ser Ser Asn Gly Ser Leu Thr Pro Ile Pro Glu Glu Arg Ser Gly
                125                   130                   135
Ser Asn Ile Cys Cys Ser Pro Ile Ser Ala Val Leu Gln Pro Leu
                140                   145                   150
Leu Leu Ile Pro Arg Gln Thr Gly Ser Gly Val Asp Leu Arg Gln
                155                   160                   165
Thr Pro Thr Arg Leu Gln Leu Arg Val Leu Thr Val Arg Arg Lys
                170                   175                   180
Thr Asn Lys Gln Lys Gly His Pro Gln Pro Tyr Pro Ile Cys Thr
                185                   190                   195
Ser Pro Ser Ser Lys Asn Lys Glu Lys Ser Leu Lys Asp Leu Met
                200                   205                   210
Glu Leu Lys Thr Thr Ala Arg Glu Leu His Asp Glu Cys Thr Ser
                215                   220                   225
Leu Ser Arg Arg Phe Asp Gln Leu Glu Glu Arg Val Ser Val Met
                230                   235                   240
Glu Asp Gln Met Ser Glu Met Lys Ile Lys Tyr Leu Gly Ile Gln
                245                   250                   255
Leu Thr Arg Asp Val Lys Glu Leu Phe Lys Glu Asn Tyr Lys Pro
                260                   265                   270
Leu Leu Asn Glu Ile Lys Glu Asp Thr Asn Lys Trp Lys Asn Ile
                275                   280                   285
Pro Cys Ser Cys Val Glu Arg Val Asn Ile Val Lys Met Ala Ile
```

```
                   290                    295                    300
Leu Pro Lys Val Ile Tyr Arg Phe Asn Ala Ile Thr Ile Lys Leu
                   305                    310                    315
Gln Ile Thr Phe Phe Thr Glu Leu Glu Lys Thr Thr Leu Lys Leu
                   320                    325                    330
Ile Trp Asn Gln Lys Arg Ala His Ile Ala Lys Ser Ile Leu Ser
                   335                    340                    345
Gln Lys Asn Lys Ala Gly Gly Ile Thr Leu Pro Asp Phe Lys Leu
                   350                    355                    360
Tyr Tyr Lys Thr Thr Val Thr Lys Thr Ala Trp Tyr Trp Tyr Gln
                   365                    370                    375
Asn Arg Gly Thr Asp Gln Trp Asn Arg Thr Glu Pro Ser Glu Ile
                   380                    385                    390
Met Leu His Ile Tyr Asn Tyr Leu Ile Phe Asp Lys Pro Asp Lys
                   395                    400                    405
Asn Lys Lys Cys Gly Lys Asp Ser Leu Phe Asn Lys Trp Cys Trp
                   410                    415                    420
Glu Asn Trp Leu Ala Ile Trp Arg Lys Leu Lys Leu Asp Pro Phe
                   425                    430                    435
Leu Thr Pro Tyr Thr Lys Ile Asn Ser Arg Trp Ile Lys Asp Leu
                   440                    445                    450
Asn Ile Arg Pro Thr Thr Ile Lys Thr Leu Glu Glu Asn Leu Gly
                   455                    460                    465
Asn Thr Ile Gln Asp Ile Gly Met Gly Lys Asn Phe Met Ser Lys
                   470                    475                    480
Thr Pro Lys Ala Val Ala Thr Lys Ala Lys Ile Tyr Lys Trp Asp
                   485                    490                    495
Leu Asn Lys Leu Lys Ser Phe Cys Thr Ala Lys Glu Ile Thr Ile
                   500                    505                    510
Arg Val Asn Arg Gln Pro Thr Lys Trp Glu Lys Ile Phe Ala Ile
                   515                    520                    525
Tyr Ser Ser Asp Lys Gly Leu Thr Ser Arg Ile Cys Asn Glu Leu
                   530                    535                    540
Lys His Ile Tyr Lys Lys Lys Thr Asn Asn Pro Ile Lys Lys Trp
                   545                    550                    555
Val Lys Asp Met Asn Arg His Phe Ser Lys Glu Asp Ile Tyr Ala
                   560                    565                    570
Val Lys Arg His Glu Lys Met Phe Ile Ile Thr Gly His Gln Arg
                   575                    580                    585
Asn Ala Asn Gln Asn His Asn Glu Lys Thr Trp Thr Gln Glu Gly
                   590                    595                    600
Glu His Asp Thr Pro Gly Pro Val Val Gly Trp Gly Glu Gly Gly
                   605                    610                    615
Gly Ile Ala Phe Gly Asp Arg Gly Pro Val Gly Val Cys Ser Tyr
                   620                    625                    630
Thr Pro Thr Pro Val Gly Arg Thr Met Ser Leu Val Ser Gln Asn
                   635                    640                    645
Ser Arg Arg Arg Arg Arg Val Ala Lys Ala Thr Ala His Asn
                   650                    655                    660
Ser Ser Trp Gly Glu Met Gln Ala Pro Asn Ala Pro Gly Leu Pro
                   665                    670                    675
Ala Asp Val Pro Gly Ser Asp Val Pro Gln Gly Pro Ser Asp Ser
                   680                    685                    690
Gln Ile Leu Gln Gly Leu Cys Ala Ser Glu Gly Pro Ser Thr Ser
                   695                    700                    705
Val Leu Pro Thr Ser Ala Glu Gly Pro Ser Thr Phe Val Pro Pro
                   710                    715                    720
Thr Ile Ser Glu Ala Ser Ser Ala Ser Gly Gln Pro Thr Ile Ser
                   725                    730                    735
Glu Gly Pro Gly Thr Ser Val Leu Pro Thr Pro Ser Glu Gly Leu
                   740                    745                    750
Ser Thr Ser Gly Pro Pro Thr Ile Ser Lys Gly Leu Cys Thr Ser
                   755                    760                    765
```

31/71

```
Val Thr Leu Ala Ala Ser Glu Gly Arg Asn Thr Ser Arg Pro Pro
            770                 775                 780
Thr Ser Ser Glu Glu Pro Ser Thr Ser Val Pro Pro Thr Ala Ser
            785                 790                 795
Glu Val Pro Ser Thr Ser Leu Pro Pro Thr Pro Gly Glu Gly Thr
            800                 805                 810
Ser Thr Ser Val Pro Pro Thr Ala Tyr Glu Gly Pro Ser Thr Ser
            815                 820                 825
Val Val Pro Thr Pro Asp Glu Gly Pro Ser Thr Ser Val Leu Pro
            830                 835                 840
Thr Pro Gly Glu Gly Pro Gly Thr Ser Val Pro Leu Ala Ala Thr
            845                 850                 855
Glu Gly Leu Ser Thr Ser Val Gln Ala Thr Pro Asp Glu Gly Pro
            860                 865                 870
Ser Thr Ser Val Pro Pro Thr Ala Thr Glu Gly Leu Ser Thr Pro
            875                 880                 885
Val Pro Pro Thr Arg Asp Glu Gly Pro Ser Thr Ser Val Pro Ala
            890                 895                 900
Thr Pro Gly Glu Gly Pro Ser Thr Ser Val Leu Pro Ala Ala Ser
            905                 910                 915
Asp Gly Gln Ser Ile Ser Leu Val Pro Thr Arg Gly Lys Gly Ser
            920                 925                 930
Ser Thr Ser Val Pro Pro Thr Ala Thr Glu Gly Leu Ser Thr Ser
            935                 940                 945
Val Gln Pro Thr Ala Gly Glu Gly Ser Ser Thr Ser Val Pro Pro
            950                 955                 960
Thr Pro Gly Gly Gly Leu Ser Thr Ser Val Pro Pro Thr Ala Thr
            965                 970                 975
Glu Glu Leu Ser Thr Ser Val Pro Pro Thr Pro Gly Glu Gly Pro
            980                 985                 990
Ser Thr Ser Val Leu Pro Ile Pro Gly Glu Gly Leu Ser Thr Ser
            995                 1000                1005
Val Pro Pro Thr Ala Ser Asp Gly Ser Asp Thr Ser Val Pro Pro
            1010                1015                1020
Thr Pro Gly Glu Gly Ala Ser Thr Leu Val Gln Pro Thr Ala Pro
            1025                1030                1035
Asp Gly Pro Gly Ser Ser Val Leu Pro Asn Pro Gly Glu Gly Pro
            1040                1045                1050
Ser Thr Leu Phe Ser Ser Ser Ala Ser Val Asp Arg Asn Pro Ser
            1055                1060                1065
Lys Cys Ser Leu Val Leu Pro Ser Pro Arg Val Thr Lys Ala Ser
            1070                1075                1080
Val Asp Ser Asp Ser Glu Gly Pro Lys Gly Ala Glu Gly Pro Ile
            1085                1090                1095
Glu Phe Glu Val Leu Arg Asp Cys Glu Ser Pro Asn Ser Ile Ser
            1100                1105                1110
Ile Met Gly Leu Asn Thr Ser Arg Val Ala Ile Thr Leu Lys Pro
            1115                1120                1125
Gln Asp Pro Met Glu Gln Asn Val Ala Glu Leu Leu Gln Phe Leu
            1130                1135                1140
Leu Val Lys Asp Gln Ser Lys Tyr Pro Ile Arg Glu Ser Glu Met
            1145                1150                1155
Arg Glu Tyr Ile Val Lys Glu Tyr Arg Asn Gln Phe Pro Glu Ile
            1160                1165                1170
Leu Arg Arg Ala Ala Ala His Leu Glu Cys Ile Phe Arg Phe Glu
            1175                1180                1185
Leu Arg Glu Leu Asp Pro Glu Ala His Thr Tyr Ile Leu Leu Asn
            1190                1195                1200
Lys Leu Gly Pro Val Pro Phe Glu Gly Leu Glu Glu Ser Pro Asn
            1205                1210                1215
Gly Pro Lys Met Gly Leu Leu Met Met Ile Leu Gly Gln Ile Phe
            1220                1225                1230
Leu Asn Gly Asn Gln Ala Lys Glu Ala Glu Ile Trp Glu Met Leu
```

```
                 1235                1240                1245
Trp Arg Met Gly Val Gln Arg Glu Arg Arg Leu Ser Ile Phe Gly
             1250                1255                1260
Asn Pro Lys Arg Leu Leu Ser Val Glu Phe Val Trp Gln Arg Tyr
             1265                1270                1275
Leu Asp Tyr Arg Pro Val Thr Asp Cys Lys Pro Val Glu Tyr Glu
             1280                1285                1290
Phe Phe Trp Gly Pro Arg Ser His Leu Glu Thr Thr Lys Met Lys
             1295                1300                1305
Ile Leu Lys Phe Met Ala Lys Ile Tyr Asn Lys Asp Pro Met Asp
             1310                1315                1320
Trp Pro Glu Lys Tyr Asn Glu Ala Leu Glu Glu Asp Ala Ala Arg
             1325                1330                1335
Ala Phe Ala Glu Gly Trp Gln Ala Leu Pro His Phe Arg Arg Pro
             1340                1345                1350
Phe Phe Glu Glu Ala Ala Ala Glu Val Pro Ser Pro Asp Ser Glu
             1355                1360                1365
Val Ser Ser Tyr Ser Ser Lys Tyr Ala Pro His Ser Trp Pro Glu
             1370                1375                1380
Ser Arg Leu Glu Ser Lys Ala Arg Lys Leu Val Gln Leu Phe Leu
             1385                1390                1395
Leu Met Asp Ser Thr Lys Leu Pro Ile Pro Lys Lys Gly Ile Leu
             1400                1405                1410
Tyr Tyr Ile Gly Arg Glu Cys Ser Lys Val Phe Pro Asp Leu Leu
             1415                1420                1425
Asn Arg Ala Ala Arg Thr Leu Asn His Val Tyr Gly Thr Glu Leu
             1430                1435                1440
Val Val Leu Asp Pro Arg Asn His Ser Tyr Thr Leu Tyr Asn Arg
             1445                1450                1455
Arg Glu Met Glu Glu Thr Glu Glu Ile Val Asp Ser Pro Asn Arg
             1460                1465                1470
Pro Gly Asn Asn Phe Leu Met Gln Val Leu Ser Phe Ile Phe Ile
             1475                1480                1485
Met Gly Asn His Ala Arg Glu Ser Ala Val Trp Ala Phe Leu Arg
             1490                1495                1500
Gly Leu Gly Val Gln Ala Gly Arg Lys His Val Ile Thr Cys Arg
             1505                1510                1515
Tyr Leu Ser Gln Arg Tyr Ile Asp Ser Leu Arg Val Pro Asp Ser
             1520                1525                1530
Asp Pro Val Gln Tyr Glu Phe Val Trp Gly Pro Arg Ala Arg Leu
             1535                1540                1545
Glu Thr Ser Lys Met Lys Ala Leu Arg Tyr Val Ala Arg Ile His
             1550                1555                1560
Arg Lys Glu Pro Gln Asp Trp Pro Gln Gln Tyr Arg Glu Ala Met
             1565                1570                1575
Glu Asp Glu Ala Asn Arg Ala Asp Val Gly His Arg Gln Ile Phe
             1580                1585                1590
Val His Asn Phe Arg
             1595
```

<210> 19
<211> 592
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6803363CD1

<400> 19
```
Met Thr Cys Trp Leu Cys Val Leu Ser Leu Pro Leu Leu Leu Leu
  1               5                  10                  15
Pro Ala Ala Pro Pro Pro Ala Gly Gly Cys Pro Ala Arg Cys Glu
```

```
                    20                    25                    30
Cys Thr Val Gln Thr Arg Ala Val Ala Cys Thr Arg Arg Arg Leu
                    35                    40                    45
Thr Ala Val Pro Asp Gly Ile Pro Ala Glu Thr Arg Leu Leu Glu
                    50                    55                    60
Leu Ser Arg Asn Arg Ile Arg Cys Leu Asn Pro Gly Asp Leu Ala
                    65                    70                    75
Ala Leu Pro Ala Leu Glu Glu Leu Asp Leu Ser Glu Asn Ala Ile
                    80                    85                    90
Ala His Val Glu Pro Gly Ala Phe Ala Asn Leu Pro Arg Leu Arg
                    95                   100                   105
Val Leu Arg Leu Arg Gly Asn Gln Leu Lys Leu Ile Pro Pro Gly
                   110                   115                   120
Val Phe Thr Arg Leu Asp Asn Leu Thr Leu Leu Asp Leu Ser Glu
                   125                   130                   135
Asn Lys Leu Val Ile Leu Leu Asp Tyr Thr Phe Gln Asp Leu His
                   140                   145                   150
Ser Leu Arg Arg Leu Glu Val Gly Asp Asn Asp Leu Val Phe Val
                   155                   160                   165
Ser Arg Arg Ala Phe Ala Gly Leu Leu Ala Leu Glu Glu Leu Thr
                   170                   175                   180
Leu Glu Arg Cys Asn Leu Thr Ala Leu Ser Gly Glu Ser Leu Gly
                   185                   190                   195
His Leu Arg Ser Leu Gly Ala Leu Arg Leu Arg His Leu Ala Ile
                   200                   205                   210
Ala Ser Leu Glu Asp Gln Asn Phe Arg Arg Leu Pro Gly Leu Leu
                   215                   220                   225
His Leu Glu Ile Asp Asn Trp Pro Leu Leu Glu Glu Val Ala Ala
                   230                   235                   240
Gly Ser Leu Arg Gly Leu Asn Leu Thr Ser Leu Ser Val Thr His
                   245                   250                   255
Thr Asn Ile Thr Ala Val Pro Ala Ala Ala Leu Arg His Gln Ala
                   260                   265                   270
His Leu Thr Cys Leu Asn Leu Ser His Asn Pro Ile Ser Thr Val
                   275                   280                   285
Pro Arg Gly Ser Phe Arg Asp Leu Val Arg Leu Arg Glu Leu His
                   290                   295                   300
Leu Ala Gly Ala Leu Leu Ala Val Val Glu Pro Gln Ala Phe Leu
                   305                   310                   315
Gly Leu Arg Gln Ile Arg Leu Leu Asn Leu Ser Asn Asn Leu Leu
                   320                   325                   330
Ser Thr Leu Glu Glu Ser Thr Phe His Ser Val Asn Thr Leu Glu
                   335                   340                   345
Thr Leu Arg Val Asp Gly Asn Pro Leu Ala Cys Asp Cys Arg Leu
                   350                   355                   360
Leu Trp Ile Val Gln Arg Arg Lys Thr Leu Asn Phe Asp Gly Arg
                   365                   370                   375
Leu Pro Ala Cys Ala Thr Pro Ala Glu Val Arg Gly Asp Ala Leu
                   380                   385                   390
Arg Asn Leu Pro Asp Ser Val Leu Phe Glu Tyr Phe Val Cys Arg
                   395                   400                   405
Lys Pro Lys Ile Arg Glu Arg Arg Leu Gln Arg Val Thr Ala Thr
                   410                   415                   420
Ala Gly Glu Asp Val Arg Phe Leu Cys Arg Ala Glu Gly Glu Pro
                   425                   430                   435
Ala Pro Thr Val Ala Trp Val Thr Pro Gln His Arg Pro Val Thr
                   440                   445                   450
Ala Thr Ser Ala Gly Arg Ala Arg Val Leu Pro Gly Gly Thr Leu
                   455                   460                   465
Glu Ile Gln Asp Ala Arg Pro Gln Asp Ser Gly Thr Tyr Thr Cys
                   470                   475                   480
Val Ala Ser Asn Ala Gly Gly Asn Asp Thr Tyr Phe Ala Thr Leu
                   485                   490                   495
```

34/71

```
Thr Val Arg Pro Glu Pro Ala Ala Asn Arg Thr Pro Gly Glu Ala
                500             505             510
His Asn Glu Thr Leu Ala Ala Leu Arg Ala Pro Leu Asp Leu Thr
                515             520             525
Thr Ile Leu Val Ser Thr Ala Met Gly Cys Ile Thr Phe Leu Gly
                530             535             540
Val Val Leu Phe Cys Phe Val Leu Leu Phe Val Trp Ser Arg Gly
                545             550             555
Arg Gly Gln His Lys Asn Asn Phe Ser Gly Glu Tyr Ser Phe Arg
                560             565             570
Lys Val Asp Gly Pro Ala Ala Ala Gly Gln Gly Gly Ala Arg
                575             580             585
Lys Phe Asn Met Lys Met Ile
                590
```

<210> 20
<211> 678
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7070580CD1

<400> 20
```
Met Val Arg Ala Gly Gly Arg Ala Trp Gln Gln Gly Leu His Ser
 1               5              10              15
Leu Ser Ala Val Ser Ile Leu Ala Lys Gly Phe Met Phe Leu Leu
                20              25              30
Val Gly Asp Ser Arg Ala Glu Ser Val Thr Asn Thr Leu Val Leu
                35              40              45
Phe Thr Ile Gly Glu Glu Ser Leu Thr Ile Phe Val Asp Lys Gln
                50              55              60
Lys Leu Gly Arg Lys Thr Glu Thr Thr Gly Gly Ala Ser Ile Ile
                65              70              75
Gly Gly Ser Gly Asn Ser Thr Ala Val Ser Leu Glu Thr Leu His
                80              85              90
Gln Leu Ala Ala Ser Tyr Phe Ile Asp Arg Glu Ser Thr Leu Arg
                95              100             105
Arg Leu His His Ile Gln Ile Ala Thr Gly Ala Ile Lys Val Thr
                110             115             120
Glu Thr Arg Thr Gly Pro Leu Gly Cys Ser Asn Tyr Asp Asn Leu
                125             130             135
Asp Ser Val Ser Ser Val Leu Val Gln Ser Pro Glu Asn Lys Val
                140             145             150
Gln Leu Leu Gly Leu Gln Val Leu Leu Pro Glu Tyr Leu Arg Glu
                155             160             165
Arg Phe Val Ala Ala Ala Leu Ser Tyr Ile Thr Cys Ser Ser Glu
                170             175             180
Gly Glu Leu Val Cys Lys Glu Asn Asp Cys Trp Cys Lys Cys Ser
                185             190             195
Pro Thr Phe Pro Glu Cys Asn Cys Pro Asp Ala Asp Ile Gln Ala
                200             205             210
Met Glu Asp Ser Leu Leu Gln Ile Gln Asp Ser Trp Ala Thr His
                215             220             225
Asn Arg Gln Phe Glu Glu Ser Glu Glu Phe Gln Ala Leu Leu Lys
                230             235             240
Arg Leu Pro Asp Asp Arg Phe Leu Asn Ser Thr Ala Ile Ser Gln
                245             250             255
Phe Trp Ala Met Asp Thr Ser Leu Gln His Arg Tyr Gln Gln Leu
                260             265             270
Gly Ala Gly Leu Lys Val Leu Phe Lys Lys Thr His Arg Ile Leu
                275             280             285
```

Arg Arg Leu Phe Asn Leu Cys Lys Arg Cys His Arg Gln Pro Arg
            290   295   300
Phe Arg Leu Pro Lys Glu Arg Ser Leu Ser Tyr Trp Trp Asn Arg
            305   310   315
Ile Gln Ser Leu Leu Tyr Cys Gly Glu Ser Thr Phe Pro Gly Thr
            320   325   330
Phe Leu Glu Gln Ser His Ser Cys Thr Cys Pro Tyr Asp Gln Ser
            335   340   345
Ser Cys Gln Gly Pro Ile Pro Cys Ala Leu Gly Glu Gly Pro Ala
            350   355   360
Cys Ala His Cys Ala Pro Asp Asn Ser Thr Arg Cys Gly Ser Cys
            365   370   375
Asn Pro Gly Tyr Val Leu Ala Gln Gly Leu Cys Arg Pro Glu Val
            380   385   390
Ala Glu Ser Leu Glu Asn Phe Leu Gly Leu Glu Thr Asp Leu Gln
            395   400   405
Asp Leu Glu Leu Lys Tyr Leu Leu Gln Lys Gln Asp Ser Arg Ile
            410   415   420
Glu Val His Ser Ile Phe Ile Ser Asn Asp Met Arg Leu Gly Ser
            425   430   435
Trp Phe Asp Pro Ser Trp Arg Lys Arg Met Leu Leu Thr Leu Lys
            440   445   450
Ser Asn Lys Tyr Lys Pro Gly Leu Val His Val Met Leu Ala Leu
            455   460   465
Ser Leu Gln Ile Cys Leu Thr Lys Asn Ser Thr Leu Glu Pro Val
            470   475   480
Met Ala Ile Tyr Val Asn Pro Phe Gly Gly Ser His Ser Glu Ser
            485   490   495
Trp Phe Met Pro Val Asn Glu Gly Ser Phe Pro Asp Trp Glu Arg
            500   505   510
Thr Asn Val Asp Ala Ala Ala Gln Cys Gln Asn Trp Thr Ile Thr
            515   520   525
Leu Gly Asn Arg Trp Lys Thr Phe Phe Glu Thr Val His Val Tyr
            530   535   540
Leu Arg Ser Arg Ile Lys Ser Leu Asp Asp Ser Ser Asn Glu Thr
            545   550   555
Ile Tyr Tyr Glu Pro Leu Glu Met Thr Asp Pro Ser Lys Asn Leu
            560   565   570
Gly Tyr Met Lys Ile Asn Thr Leu Gln Val Phe Gly Tyr Ser Leu
            575   580   585
Pro Phe Asp Pro Asp Ala Ile Arg Asp Leu Ile Leu Gln Phe Asp
            590   595   600
Tyr Pro Tyr Thr Gln Gly Ser Gln Asp Ser Ala Leu Leu Gln Leu
            605   610   615
Ile Glu Leu Arg Asp Arg Val Asn Gln Leu Ser Pro Pro Gly Lys
            620   625   630
Val Arg Leu Asp Leu Phe Ser Cys Leu Leu Arg His Arg Leu Lys
            635   640   645
Leu Ala Asn Asn Glu Val Gly Arg Ile Gln Ser Ser Leu Arg Ala
            650   655   660
Phe Asn Ser Lys Leu Pro Asn Pro Val Glu Tyr Glu Thr Gly Lys
            665   670   675
Leu Cys Ser

<210> 21
<211> 652
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500176CD1

36/71

<400> 21

```
Met Ala Ala Glu Thr Leu Leu Ser Ser Leu Leu Gly Leu Leu Leu
1               5                   10                  15
Leu Gly Leu Leu Leu Pro Ala Ser Leu Thr Gly Gly Val Gly Ser
            20                  25                  30
Leu Asn Leu Glu Glu Leu Ser Glu Met Arg Tyr Gly Ile Glu Ile
            35                  40                  45
Leu Pro Leu Pro Val Met Gly Gly Gln Ser Gln Ser Ser Asp Val
            50                  55                  60
Val Ile Val Ser Ser Lys Tyr Lys Gln Arg Tyr Glu Cys Arg Leu
            65                  70                  75
Pro Ala Gly Ala Ile His Phe Gln Arg Glu Arg Glu Glu Glu Thr
            80                  85                  90
Pro Ala Tyr Gln Gly Pro Gly Ile Pro Glu Leu Leu Ser Pro Met
            95                  100                 105
Arg Asp Ala Pro Cys Leu Leu Lys Thr Lys Asp Trp Trp Thr Tyr
            110                 115                 120
Glu Phe Cys Tyr Gly Arg His Ile Gln Gln Tyr His Met Glu Asp
            125                 130                 135
Ser Glu Ile Lys Gly Glu Val Leu Tyr Leu Gly Tyr Tyr Gln Ser
            140                 145                 150
Ala Phe Asp Trp Asp Asp Glu Thr Ala Lys Ala Ser Lys Gln His
.           155                 160                 165
Arg Leu Lys Arg Tyr His Ser Gln Thr Tyr Gly Asn Gly Ser Lys
            170                 175                 .   180
Cys Asp Leu Asn Gly Arg Pro Arg Glu Ala Glu Val Arg Phe Leu
            185                 190                 195
Cys Asp Glu Gly Ala Gly Ile Ser Gly Asp Tyr Ile Asp Arg Val
.           200                 205                 210
Asp Glu Pro Leu Ser Cys Ser Tyr Val Leu Thr Ile Arg Thr Pro
            215                 220                 225
Arg Leu Cys Pro His Pro Leu Leu Arg Pro Pro Ser Ala Ala
            230                 235                 240
Pro Gln Ala Ile Leu Cys His Pro Ser Leu Gln Pro Glu Glu Tyr
            245                 250                 255
Met Ala Tyr Val Gln Arg Gln Ala Asp Ser Lys Gln Tyr Gly Asp
            260                 265                 270
Lys Ile Ile Glu Glu Leu Gln Asp Leu Gly Pro Gln Val Trp Ser
            275                 280                 285
Glu Thr Lys Ser Gly Val Ala Pro Gln Lys Met Ala Gly Ala Ser
            290                 295                 300
Pro Thr Lys Asp Asp Ser Lys Asp Ser Asp Phe Trp Lys Met Leu
            305                 310                 315
Asn Glu Pro Glu Asp Gln Ala Pro Gly Gly Glu Glu Val Pro Ala
            320                 325                 330
Glu Glu Gln Asp Pro Ser Pro Glu Ala Ala Asp Ser Ala Ser Gly
            335                 340                 345
Ala Pro Asn Asp Phe Gln Asn Asn Val Gln Val Lys Val Ile Arg
.           350                 355                 360
Ser Pro Ala Asp Leu Ile Arg Phe Ile Glu Glu Leu Lys Gly Gly
            365                 370                 375
Thr Lys Lys Gly Lys Pro Asn Ile Gly Gln Glu Gln Pro Val Asp
            380                 385                 390
Asp Ala Ala Glu Val Pro Gln Arg Glu Pro Glu Lys Glu Arg Gly
            395                 400                 405
Asp Pro Glu Arg Gln Arg Glu Met Glu Glu Glu Asp Glu Asp
            410                 415                 420
Glu Asp Glu Asp Glu Asp Glu Asp Glu Arg Gln Leu Leu Gly Glu
            425                 430                 435
Phe Glu Lys Glu Leu Glu Gly Ile Leu Leu Pro Ser Asp Arg Asp
            440                 445                 450
Arg Leu Arg Ser Glu Thr Glu Lys Glu Leu Asp Pro Asp Gly Leu
            455                 460                 465
```

```
Lys Lys Glu Ser Glu Arg Asp Arg Ala Met Leu Ala Leu Thr Ser
                470             475             480
Thr Leu Asn Lys Leu Ile Lys Arg Leu Glu Glu Lys Gln Ser Pro
                485             490             495
Glu Leu Val Lys Lys His Lys Lys Lys Arg Val Val Pro Lys Lys
                500             505             510
Pro Pro Pro Ser Pro Gln Pro Thr Glu Glu Asp Pro Glu His Arg
                515             520             525
Val Arg Val Arg Val Thr Lys Leu Arg Leu Gly Gly Pro Asn Gln
                530             535             540
Asp Leu Thr Val Leu Glu Met Lys Arg Glu Asn Pro Gln Leu Lys
                545             550             555
Gln Ile Glu Gly Leu Val Lys Glu Leu Leu Glu Arg Glu Gly Leu
                560             565             570
Thr Ala Ala Gly Lys Ile Glu Ile Lys Ile Val Arg Pro Trp Ala
                575             580             585
Glu Gly Thr Glu Glu Gly Ala Arg Trp Leu Thr Asp Glu Asp Thr
                590             595             600
Arg Asn Leu Lys Glu Ile Phe Phe Asn Ile Leu Val Pro Gly Ala
                605             610             615
Glu Glu Ala Gln Lys Glu Arg Gln Arg Gln Lys Glu Leu Glu Ser
                620             625             630
Asn Tyr Arg Arg Val Trp Gly Ser Pro Gly Gly Glu Gly Thr Gly
                635             640             645
Asp Leu Asp Glu Phe Asp Phe
                650
```

```
<210> 22
<211> 457
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500506CD1

<400> 22
Met Ala Leu Pro Thr Ala Arg Pro Leu Leu Gly Ser Cys Gly Thr
  1           5             10              15
Pro Ala Leu Gly Ser Leu Leu Phe Leu Leu Phe Ser Leu Gly Trp
                20              25              30
Val Gln Pro Ser Arg Thr Leu Ala Gly Glu Thr Gly Gln Glu Ala
                35              40              45
Ala Pro Leu Asp Gly Val Leu Ala Asn Pro Pro Asn Ile Ser Ser
                50              55              60
Leu Ser Pro Arg Gln Leu Leu Gly Phe Pro Cys Ala Glu Val Ser
                65              70              75
Gly Leu Ser Thr Glu Arg Val Arg Glu Leu Ala Val Ala Leu Ala
                80              85              90
Gln Lys Asn Val Lys Leu Ser Thr Glu Gln Gly Ile Val Ala Ala
                95              100             105
Trp Arg Gln Arg Ser Ser Arg Asp Pro Ser Trp Arg Gln Pro Glu
                110             115             120
Arg Thr Ile Leu Arg Pro Arg Phe Arg Arg Glu Val Glu Lys Thr
                125             130             135
Ala Cys Pro Ser Gly Lys Lys Ala Arg Glu Ile Asp Glu Ser Leu
                140             145             150
Ile Phe Tyr Lys Lys Trp Glu Leu Glu Ala Cys Val Asp Ala Ala
                155             160             165
Leu Leu Ala Thr Gln Met Asp Arg Val Asn Ala Ile Pro Phe Thr
                170             175             180
Tyr Glu Gln Leu Asp Val Leu Lys His Lys Leu Asp Glu Leu Tyr
                185             190             195
```

```
Pro Gln Gly Tyr Pro Glu Ser Val Ile Gln His Leu Gly Tyr Leu
            200                   205                   210
Phe Leu Lys Met Ser Pro Glu Asp Ile Arg Lys Trp Asn Val Thr
            215                   220                   225
Ser Leu Glu Thr Leu Lys Ala Leu Leu Glu Val Asn Lys Gly His
            230                   235                   240
Glu Met Ser Pro Gln Val Ala Thr Leu Ile Asp Arg Phe Val Lys
            245                   250                   255
Gly Arg Gly Gln Leu Asp Lys Asp Thr Leu Asp Thr Leu Thr Ala
            260                   265                   270
Phe Tyr Pro Gly Tyr Leu Cys Ser Leu Ser Pro Glu Glu Leu Ser
            275                   280                   285
Ser Val Pro Pro Ser Ser Ile Trp Ala Val Arg Pro Gln Asp Leu
            290                   295                   300
Asp Thr Cys Asp Pro Arg Gln Leu Asp Val Leu Tyr Pro Lys Ala
            305                   310                   315
Arg Leu Ala Phe Gln Asn Met Asn Gly Ser Glu Tyr Phe Val Lys
            320                   325                   330
Ile Gln Ser Phe Leu Gly Gly Ala Pro Thr Glu Asp Leu Lys Ala
            335                   340                   345
Leu Ser Gln Gln Asn Val Ser Met Asp Leu Ala Thr Phe Met Lys
            350                   355                   360
Leu Arg Thr Asp Ala Val Leu Pro Leu Thr Val Ala Glu Val Gln
            365                   370                   375
Lys Leu Leu Gly Pro His Val Glu Gly Leu Lys Ala Glu Glu Arg
            380                   385                   390
His Arg Pro Val Arg Asp Trp Ile Leu Arg Gln Arg Gln Asp Asp
            395                   400                   405
Leu Asp Thr Leu Gly Leu Gly Leu Gln Gly Gly Ile Pro Asn Gly
            410                   415                   420
Tyr Leu Val Leu Asp Leu Ser Val Gln Glu Ala Leu Ser Gly Thr
            425                   430                   435
Pro Cys Leu Leu Gly Pro Gly Pro Val Leu Thr Val Leu Ala Leu
            440                   445                   450
Leu Leu Ala Ser Thr Leu Ala
            455
```

```
<210> 23
<211> 127
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500639CD1

<400> 23
Met Glu Thr Ala Val Arg Gly Met Pro Leu Glu Cys Pro Pro Arg
  1           5                  10                    15
Pro Glu Arg Leu Asn Ala Tyr Glu Arg Glu Val Met Val Asn Met
            20                  25                    30
Leu Asn Ser Leu Ser Arg Asn Gln Gln Leu Pro Arg Ile Thr Pro
            35                  40                    45
Arg Cys Gly Cys Val Asp Pro Leu Pro Gly Arg Leu Pro Phe His
            50                  55                    60
Gly Tyr Glu Ser Ala Cys Ser Gly Arg His Tyr Cys Leu Arg Gly
            65                  70                    75
Met Asp Tyr Tyr Ala Ser Gly Ala Pro Cys Thr Asp Arg Arg Leu
            80                  85                    90
Arg Pro Trp Cys Arg Glu Gln Pro Thr Met Cys Thr Ser Leu Arg
            95                  100                   105
Ala Pro Ala Arg Asn Ala Val Cys Cys Tyr Asn Ser Pro Ala Val
            110                 115                   120
```

39/71

Ile Leu Pro Ile Ser Glu Pro
                125

<210> 24
<211> 763
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506167CD1

<400> 24
Met Glu Glu Pro Gly Ala Thr Pro Gln Pro Tyr Leu Gly Leu Val
1   5   10   15
Leu Glu Glu Leu Arg Arg Val Val Ala Ala Leu Pro Glu Ser Met
                20   25   30
Arg Pro Asp Glu Asn Pro Tyr Gly Phe Pro Ser Glu Leu Val Val
                35   40   45
Cys Ala Ala Val Ile Gly Phe Phe Val Val Leu Leu Phe Leu Trp
                50   55   60
Arg Ser Phe Arg Ser Val Arg Ser Arg Leu Tyr Val Gly Arg Glu
                65   70   75
Gln Lys Leu Gly Ala Thr Leu Ser Gly Leu Ile Glu Glu Lys Cys
                80   85   90
Lys Leu Leu Glu Lys Phe Ser Leu Ile Gln Lys Glu Tyr Glu Gly
                95   100   105
Tyr Glu Val Glu Ser Ser Leu Glu Asp Ala Ser Phe Glu Lys Ala
                110   115   120
Ala Ala Glu Glu Ala Arg Ser Leu Glu Ala Thr Cys Glu Lys Leu
                125   130   135
Asn Arg Ser Asn Ser Glu Leu Glu Asp Glu Ile Leu Cys Leu Glu
                140   145   150
Lys Asp Leu Lys Gln Glu Lys Ser Lys His Ser Gln Gln Asp Glu
                155   160   165
Leu Met Ala Asp Ile Ser Lys Ser Ile Gln Ser Leu Glu Asp Glu
                170   175   180
Ser Lys Ser Leu Lys Ser Gln Ile Ala Glu Ala Lys Ile Ile Cys
                185   190   195
Lys Thr Phe Lys Met Ser Glu Glu Arg Arg Ala Ile Ala Ile Lys
                200   205   210
Asp Ala Leu Asn Glu Asn Ser Gln Leu Gln Thr Ser His Lys Gln
                215   220   225
Leu Phe Gln Gln Glu Ala Glu Val Trp Lys Gly Glu Val Ser Glu
                230   235   240
Leu Asn Lys Gln Lys Ile Thr Phe Glu Asp Ser Lys Val His Ala
                245   250   255
Glu Gln Val Leu Asn Asp Lys Glu Asn His Ile Lys Thr Leu Thr
                260   265   270
Gly His Leu Pro Met Met Lys Asp Gln Ala Ala Val Leu Glu Glu
                275   280   285
Asp Thr Thr Asp Asp Asp Asn Leu Glu Leu Glu Val Asn Ser Gln
                290   295   300
Trp Glu Asn Gly Ala Tyr Leu Asp Asn Pro Pro Lys Gly Ala Leu
                305   310   315
Lys Lys Leu Ile His Ala Ala Lys Leu Asn Ala Ser Leu Lys Thr
                320   325   330
Leu Glu Gly Glu Arg Asn Gln Ile Tyr Ile Gln Leu Ser Glu Val
                335   340   345
Asp Lys Thr Lys Glu Glu Leu Thr Glu His Ile Lys Asn Leu Gln
                350   355   360
Thr Gln Gln Ala Ser Leu Gln Ser Glu Asn Thr His Phe Glu Asn
                365   370   375

40/71

Glu Asn Gln Lys Leu Gln Gln Lys Leu Lys Val Met Thr Glu Leu
            380   385   390
Tyr Gln Glu Asn Glu Met Lys Leu His Arg Lys Leu Thr Val Glu
            395   400   405
Glu Asn Tyr Arg Leu Glu Lys Glu Glu Lys Leu Ser Lys Val Asp
            410   415   420
Glu Lys Ile Ser His Ala Thr Glu Glu Leu Glu Thr Tyr Arg Lys
            425   430   435
Arg Ala Lys Asp Leu Glu Glu Glu Leu Glu Arg Thr Ile His Ser
            440   445   450
Tyr Gln Gly Gln Ile Ile Ser His Glu Lys Lys Ala His Asp Asn
            455   460   465
Trp Leu Ala Ala Arg Asn Ala Glu Arg Asn Leu Asn Asp Leu Arg
            470   475   480
Lys Glu Asn Ala His Asn Arg Gln Lys Leu Thr Glu Thr Glu Leu
            485   490   495
Lys Phe Glu Leu Leu Glu Lys Asp Pro Tyr Ala Leu Asp Val Pro
            500   505   510
Asn Thr Ala Phe Gly Arg Gly Ser Arg Gly Pro Gly Asn Pro Leu
            515   520   525
Asp His Gln Ile Thr Asn Glu Arg Gly Glu Ser Ser Cys Asp Arg
            530   535   540
Leu Thr Asp Pro His Arg Ala Pro Ser Asp Thr Gly Ser Leu Ser
            545   550   555
Pro Pro Trp Asp Gln Asp Arg Arg Met Met Phe Pro Pro Pro Gly
            560   565   570
Gln Ser Tyr Pro Asp Ser Ala Leu Pro Pro Gln Arg Gln Asp Arg
            575   580   585
Phe Cys Ser Asn Ser Gly Arg Leu Ser Gly Pro Ala Glu Leu Arg
            590   595   600
Ser Phe Asn Met Pro Ser Leu Asp Lys Met Asp Gly Ser Met Pro
            605   610   615
Ser Glu Met Glu Ser Ser Arg Asn Asp Thr Lys Asp Asp Leu Gly
            620   625   630
Asn Leu Asn Val Pro Asp Ser Ser Leu Pro Ala Glu Asn Glu Ala
            635   640   645
Thr Gly Pro Gly Phe Val Pro Pro Pro Leu Ala Pro Ile Arg Gly
            650   655   660
Pro Leu Phe Pro Val Asp Ala Arg Gly Pro Phe Leu Arg Arg Gly
            665   670   675
Pro Pro Phe Pro Pro Pro Pro Gly Ala Met Phe Gly Ala Ser
            680   685   690
Arg Asp Tyr Phe Pro Pro Arg Asp Phe Pro Gly Pro Pro Ala
            695   700   705
Pro Phe Ala Met Arg Asn Val Tyr Pro Pro Arg Gly Phe Pro Pro
            710   715   720
Tyr Leu Pro Pro Arg Pro Gly Phe Phe Pro Pro Pro Pro His Ser
            725   730   735
Glu Gly Arg Ser Glu Phe Pro Ser Gly Leu Ile Pro Pro Ser Asn
            740   745   750
Glu Pro Ala Thr Glu His Pro Glu Pro Gln Gln Glu Thr
            755   760

<210> 25
<211> 612
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 90081189CD1

<400> 25

```
Met Ala Ala Glu Thr Leu Leu Ser Ser Leu Leu Gly Leu Leu Leu
1               5                   10                  15
Leu Gly Leu Leu Leu Pro Ala Ser Leu Thr Gly Gly Val Gly Ser
            20                  25                  30
Leu Asn Leu Glu Glu Leu Ser Glu Met Arg Tyr Gly Ile Glu Ile
            35                  40                  45
Leu Pro Leu Pro Val Met Gly Gly Gln Ser Gln Ser Ser Asp Val
            50                  55                  60
Val Ile Val Ser Ser Lys Tyr Lys Gln Arg Tyr Glu Cys Arg Leu
            65                  70                  75
Pro Ala Gly Ala Ile His Phe Gln Arg Glu Arg Glu Glu Glu Thr
            80                  85                  90
Pro Ala Tyr Gln Gly Pro Gly Ile Pro Glu Leu Leu Ser Pro Met
            95                  100                 105
Arg Asp Ala Pro Cys Leu Leu Lys Thr Lys Asp Trp Trp Thr Tyr
            110                 115                 120
Glu Phe Cys Tyr Gly Arg His Ile Gln Gln Tyr His Met Glu Asp
            125                 130                 135
Ser Glu Ile Lys Gly Glu Val Leu Tyr Leu Gly Tyr Tyr Gln Ser
            140                 145                 150
Ala Phe Asp Trp Asp Asp Glu Thr Ala Lys Ala Ser Lys Gln His
            155                 160                 165
Arg Leu Lys Arg Tyr His Ser Gln Thr Tyr Gly Asn Gly Ser Lys
            170                 175                 180
Cys Asp Leu Asn Gly Arg Pro Arg Glu Ala Glu Val Arg Phe Leu
            185                 190                 195
Cys Asp Glu Gly Ala Gly Ile Ser Gly Asp Tyr Ile Asp Arg Val
            200                 205                 210
Asp Glu Pro Leu Ser Cys Ser Tyr Val Leu Thr Ile Arg Thr Pro
            215                 220                 225
Arg Leu Cys Pro His Pro Leu Leu Arg Pro Pro Pro Ser Ala Ala
            230                 235                 240
Pro Gln Ala Ile Leu Cys His Pro Ser Leu Gln Pro Glu Glu Tyr
            245                 250                 255
Met Ala Tyr Val Gln Arg Gln Ala Asp Ser Lys Gln Tyr Gly Asp
            260                 265                 270
Lys Ile Ile Glu Glu Leu Gln Asp Leu Gly Pro Gln Val Trp Ser
            275                 280                 285
Glu Thr Lys Ser Gly Val Ala Pro Gln Lys Met Ala Gly Ala Ser
            290                 295                 300
Pro Thr Lys Asp Asp Ser Lys Asp Pro Asp Phe Trp Lys Met Leu
            305                 310                 315
Asn Glu Pro Glu Asp Gln Ala Pro Gly Gly Glu Glu Val Pro Ala
            320                 325                 330
Glu Glu Gln Asp Pro Ser Pro Glu Ala Ala Asp Ser Ala Ser Gly
            335                 340                 345
Ala Pro Asn Asp Phe Gln Asn Asn Val Gln Val Lys Val Ile Arg
            350                 355                 360
Ser Pro Ala Asp Leu Ile Arg Phe Ile Glu Glu Leu Lys Gly Gly
            365                 370                 375
Thr Lys Lys Gly Lys Pro Asn Ile Gly Gln Glu Gln Pro Val Asp
            380                 385                 390
Asp Ala Ala Glu Val Pro Gln Arg Glu Pro Glu Lys Glu Arg Gly
            395                 400                 405
Asp Pro Glu Arg Gln Arg Glu Met Glu Glu Glu Asp Glu Asp
            410                 415                 420
Glu Asp Glu Asp Glu Asp Glu Asp Arg Gln Leu Leu Gly Glu
            425                 430                 435
Phe Glu Lys Glu Leu Glu Gly Ile Leu Leu Pro Ser Asp Arg Asp
            440                 445                 450
Arg Leu Arg Ser Glu Val Lys Ala Gly Met Glu Arg Glu Leu Glu
            455                 460                 465
Asn Ile Ile Gln Glu Thr Glu Lys Glu Leu Asp Pro Asp Gly Leu
```

```
                  470                    475                    480
Lys Lys Glu Ser Glu Arg Asp Arg Ala Met Leu Ala Leu Thr Ser
                  485                    490                    495
Thr Leu Asn Lys Leu Ile Lys Arg Leu Glu Glu Lys Gln Ser Pro
                  500                    505                    510
Glu Leu Val Lys Lys His Lys Lys Lys Arg Val Val Pro Lys Lys
                  515                    520                    525
Pro Pro Pro Ser Pro Gln Pro Thr Gly Lys Ile Glu Ile Lys Ile
                  530                    535                    540
Val Arg Pro Trp Ala Glu Gly Thr Glu Glu Gly Ala Arg Trp Leu
                  545                    550                    555
Thr Asp Glu Asp Thr Arg Asn Leu Lys Glu Ile Phe Phe Asn Ile
                  560                    565                    570
Leu Val Pro Gly Ala Glu Glu Ala Gln Lys Glu Arg Gln 'Arg Gln
                  575                    580                    585
Lys Glu Leu Glu Ser Asn Tyr Arg Arg Val Trp Gly Ser Pro Gly
                  590                    595                    600
Gly Glu Gly Thr Gly Asp Leu Asp Glu Phe Asp Phe
                  605                    610
```

```
<210> 26
<211> 638
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7510095CD1

<400> 26
Met Ala Ala Glu Thr Leu Leu Ser Ser Leu Leu Gly Leu Leu Leu
1               5                   10                     15
Leu Gly Leu Leu Leu Pro Ala Ser Leu Thr Gly Gly Val Gly Ser
                  20                   25                     30
Leu Asn Leu Glu Glu Leu Ser Glu Met Arg Tyr Gly Ile Glu Ile
                  35                   40                     45
Leu Pro Leu Pro Val Met Gly Gly Gln Ser Gln Ser Ser Asp Val
                  50                   55                     60
Val Ile Val Ser Ser Lys Tyr Lys Gln Arg Tyr Glu Cys Arg Leu
                  65                   70                     75
Pro Ala Gly Ala Ile His Phe Gln Arg Glu Arg Glu Glu Glu Thr
                  80                   85                     90
Pro Ala Tyr Gln Gly Pro Gly Ile Pro Glu Leu Leu Ser Pro Met
                  95                  100                    105
Arg Asp Ala Pro Cys Leu Leu Lys Thr Lys Asp Trp Trp Thr Tyr
                  110                  115                    120
Glu Phe Cys Tyr Gly Arg His Ile Gln Gln Tyr His Met Glu Asp
                  125                  130                    135
Ser Glu Ile Lys Gly Glu Val Leu Tyr Leu Gly Tyr Tyr Gln Ser
                  140                  145                    150
Ala Phe Asp Trp Asp Asp Glu Thr Ala Lys Ala Ser Lys Gln His
                  155                  160                    165
Arg Leu Lys Arg Tyr His Ser Gln Thr Tyr Gly Asn Gly Ser Lys
                  170                  175                    180
Cys Asp Leu Asn Gly Arg Pro Arg Glu Ala Glu Val Arg Phe Leu
                  185                  190                    195
Cys Asp Glu Gly Ala Gly Ile Ser Gly Asp Tyr Ile Asp Arg Val
                  200                  205                    210
Asp Glu Pro Leu Ser Cys Ser Tyr Val Leu Thr Ile Arg Thr Pro
                  215                  220                    225
Arg Leu Cys Pro His Pro Leu Leu Arg Pro Pro Ser Ala Ala
                  230                  235                    240
Pro Gln Ala Ile Leu Cys His Pro Ser Leu Gln Pro Glu Glu Tyr
```

```
                   245                     250                     255
Met Ala Tyr Val Gln Arg Gln Ala Asp Ser Lys Gln Tyr Gly Asp
                   260                     265                     270
Lys Ile Ile Glu Glu Leu Gln Asp Leu Gly Pro Gln Val Trp Ser
                   275                     280                     285
Glu Thr Lys Ser Gly Val Ala Pro Gln Lys Met Ala Gly Ala Ser
                   290                     295                     300
Pro Thr Lys Asp Asp Ser Lys Asp Ser Asp Phe Trp Lys Met Leu
                   305                     310                     315
Asn Glu Pro Glu Asp Gln Ala Pro Gly Gly Glu Glu Val Pro Ala
                   320                     325                     330
Glu Glu Gln Asp Pro Ser Pro Glu Ala Ala Asp Ser Ala Ser Gly
                   335                     340                     345
Ala Pro Asn Asp Phe Gln Asn Asn Val Gln Val Lys Val Ile Arg
                   350                     355                     360
Ser Pro Ala Asp Leu Ile Arg Phe Ile Glu Glu Leu Lys Gly Gly
                   365                     370                     375
Thr Lys Lys Gly Lys Pro Asn Ile Gly Gln Glu Gln Pro Val Asp
                   380                     385                     390
Asp Ala Ala Glu Val Pro Gln Arg Glu Pro Glu Lys Glu Arg Gly
                   395                     400                     405
Asp Pro Glu Arg Gln Arg Glu Met Glu Glu Glu Glu Asp Glu Asp
                   410                     415                     420
Glu Asp Glu Asp Glu Asp Glu Asp Glu Arg Gln Leu Leu Gly Glu
                   425                     430                     435
Phe Glu Lys Glu Leu Glu Gly Ile Leu Leu Pro Ser Asp Arg Asp
                   440                     445                     450
Arg Leu Arg Ser Glu Val Lys Ala Gly Met Glu Arg Glu Leu Glu
                   455                     460                     465
Asn Ile Ile Gln Glu Thr Glu Lys Glu Leu Asp Pro Asp Gly Leu
                   470                     475                     480
Lys Lys Glu Ser Glu Arg Asp Arg Ala Met Leu Ala Leu Thr Ser
                   485                     490                     495
Thr Leu Asn Lys Leu Ile Lys Arg Leu Glu Glu Lys Gln Ser Pro
                   500                     505                     510
Glu Leu Val Lys Lys His Lys Lys Lys Arg Val Val Pro Lys Lys
                   515                     520                     525
Pro Pro Pro Ser Pro Gln Pro Thr Glu Glu Asp Pro Glu His Arg
                   530                     535                     540
Val Arg Val Arg Val Thr Lys Leu Arg Leu Gly Gly Pro Asn Gln
                   545                     550                     555
Asp Leu Thr Val Leu Glu Met Lys Arg Glu Asn Pro Gln Leu Lys
                   560                     565                     570
Gln Ile Glu Gly Leu Val Lys Glu Leu Leu Glu Arg Glu Gly Leu
                   575                     580                     585
Thr Ala Ala Gly Gly Pro Trp Arg Ala Ala Gly Pro Ser Ala Val
                   590                     595                     600
Gly Arg Ala Ser Cys Arg Lys Trp Arg Gly Trp Asp Gln Trp Gly
                   605                     610                     615
Arg Gly Pro Ala Leu Arg Leu Pro Thr Ile Leu Ser Ser His Cys
                   620                     625                     630
Ile Leu Leu Leu Ser Arg Glu Asn
                   635
```

<210> 27
<211> 75
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7510096CD1

<400> 27
Met Ala Ala Glu Thr Leu Leu Ser Ser Leu Leu Gly Leu Leu Leu
1   5   10   15
Leu Gly Leu Leu Leu Pro Ala Ser Leu Thr Gly Gly Val Gly Ser
                20   25   30
Leu Asn Leu Glu Glu Leu Ser Glu Met Arg Tyr Gly Ile Glu Ile
                35   40   45
Leu Pro Leu Pro Val Met Gly Gly Gln Pro Phe Phe Ala Ser Leu
                50   55   60
Pro Pro Asp Pro Pro Phe Phe Pro Arg Ala Asn Leu Arg Thr Trp
                65   70   75

<210> 28
<211> 3343
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 261510CB1

<400> 28
ccctcttccg tgtgggccct ccctgtcccc tctctgtcac ctcacagtag tcgttgccgg 60
tagcgggtgt gggggcgcga gcagcgcagg gcggcctctc ccccacccct cagcccgggc 120
ccgggctcgg tcccccctc tccctccgct tcgcactccc gtcgggtacg ggaaggtcca 180
agccgctgcc gggtgccga gggccgtcgt ggccgccgtc gccacgggtc ccaatggagc 240
cgccgaatct ctatccggtg aagctctacg tgtacgacct gtccaaaggc ctggcccggc 300
ggctcagccc catcatgctg gggaaacaac tggaaggcat ctggcacaca tccatagttg 360
tgcacaagga tgagttcttc ttcggcagtg gtggtatctc cagctgcccc ccgggaggga 420
cattgcttgg gcctccagac tctgtggttg atgtggggag tacagaagtc acagaagaaa 480
tctttctgga gtacctctcc tccctggggg agtccctgtt ccgaggtgag gcctacaacc 540
tctttgaaca caattgtaac accttcagca cgaagtggc acagttcctg actgggcgga 600
agattccttc ttacatcaca gacctgccct ctgaagttct ctccacgccc tttggacagg 660
cacttcggcc cctcctggac tccattcaga tccagcctcc aggagggagc tccgtgggca 720
gacccaacgg ccagagctaa caggactgcc tgggaccgcc ctgcctcacc agggcttttc 780
ctttttaaac aaaacaaacc ctaccagatt tctattttat aattttacat cagagctaac 840
aaccagggga cggcttttta aatttcccag ggaaggagac cgtcaggccg catgtagaca 900
atgctgctaa gaaacagaac aaaatgccac cccttctaat agtattatac taatttatta 960
aggctgtctt gtctgtgcgt tcacttcgga tgctgtttcc taagcgtcct ccacactgga 1020
gaaagggagg ggctttgttt aaacagaaag cgagtcacag tttgcgggag ccccaaccta 1080
gaactctttc cagagaacac tcacgctcta cccaagttgc caacacctgt gggacccagg 1140
tgagttttag aacatccgtg acccagaaag ttactctagt gttaaatgaa ccaaaggatt 1200
tggtgtacat gggaccggtt tcaggatgac taacctaacc taagtgaact ctttagccac 1260
aaccctcacc ttccctccct taccccaagc aaggccagtg ctggagcagc cattgcctcc 1320
ggatcaaggg agtaacagca agtatcaggc cggagaaaac aggataaagc cattcctcct 1380
tccacctaga tggattgcca atcttctgcc atcaattttg ggaaacctgc ctttgtgccc 1440
ccgaggctgc ctttcctcag catagctgac ttgtcatagt ggatgaggtc tgggctggag 1500
cagtagaagc caaggccttt tttctgcaag taatgatgta ggttatttgc agtctctgta 1560
accaccatcc aggccagtac gcagtattta gctggggacc aagggatggg attccctact 1620
ctaagaaaat acctttattt taattattat tattttttga gacggagtct cgctctacca 1680
cccaggctgg tgtgcagtgg catgatcccg ggaggcagag attgcgatga gccgagatcg 1740
taccagtgca ctccagcctg ggcgacagag tgagactcca tctcaaaaaa ataaataat 1800
gaagtggaat caaggtggca agaaaattta gaatcgtttc tgatagggac tttcccttat 1860
cctaaatgtg ttttgaaggc acacgttatt tccattgtca ttctgacttt agaaaggcag 1920
gttggggcat gagtgggcat tttcctaatg acttgctgct aggaggttat aggcagggaa 1980
cagtctggga gtctctaatc cagagagact ttcagattcc ctctcttcac atgaacctcc 2040
attagttaga ttttgggaaa gtgagcatga agaagggact cagagctgcc aggtacccga 2100
gtctcgggtc tgacatcctt gggcccccttc tgccacaact tcttgccact tcacggactt 2160
gaattatgtt gcttcttcct ggaagtctgg cctgcagcac aggcctttcc tctccaaaaa 2220
gctgctgctc atttctggcc ccatctgctg ctatctttgc catcagccat cacagtttgg 2280
catcatcaga gtctcagcta gaggagctgt cattctgcct tagtgtttgc ctgcgacctg 2340
accaggggca tgaccagctg ggatgtcaag tgacatttgg gaacttgtag aggccaaatg 2400

```
attgaaagtt taattcttct caagacccac tgatttgcca gtgtgcatgg aaataataga 2460
ttagagcaga aactagcagg gactgttgta taatcgtgat ctactagcag aattgggccc 2520
aaagagaaag aatttctgag gctgggtttc ttctgggagg gtagcaggtt ggtcttggcc 2580
agtgctctgt ttcccaccag tgttgttccc ttttccagct tgcctggccc ggctctgtct 2640
tccagggtcc tgctccagac ccctctgcct ggggtctccc ctccatccag ggggttctgg 2700
gcatcacagc tcctttttt tcctttctcc tcttccagat ccagtttcag ctggtttgct 2760
tcttccagca ggcacccttg ggcccaggag cagctccct gagagaattt gagtcactca 2820
tttccttgaa cagcaccccc ctttatacta gcagccattt gtgccattgc ctgtgcccta 2880
gggtttgtgg ggagagagcg agggatcact gagcagtttt cccagagctc catgggaagg 2940
caagctctcc ctcccaatgg gagcccccact gtcactaact gtaaactcag gctcaggctt 3000
caactgccta cccccatcct catatttctg tctgtcccag cacctcagga gcattctcat 3060
tgtggccggc taactccgcc tggatgtgaa caggcaagca cagtgggaaa tgagtcacgt 3120
acttgtattg cacagtggac acctctagag gtccattggt ttaaagggat agggaaggag 3180
gagggatgag accatcaccc cctcccagaa gtaaatctag tatctgagtt ttctttatgc 3240
ccttgagtca aactaataac tgtctagtac ggaggtgttt gctggttttc ttcggtgttt 3300
tttctaatgc aataaactca tttctgcctg ctcgaaaaaa aaa                   3343
```

<210> 29
<211> 3190
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7498674CB1

<400> 29
```
cggccccagg gagagatcgc gggcagcatg gcggggctgc gggccagggg gggcccgggg 60
ccggggctgc tggcgctctc cgcgctcggc ttctgcctca tgctgcaagt cagcgctaag 120
aggcccccca agacgccccc ctgcccgccc agctgctctt gcaccaggga caccgccttc 180
tgcgtggact caaaggcggt gcccaggaac ctgccctcgg aggtcatctc cctgaccctg 240
gtgaatgccg ccttctcaga gatccaggat ggagcgttct cccacctccc gctgctgcag 300
ttcttgttac tcaactccaa caagtttaca ctgattggag acaacgcctt cacaggactg 360
tcgcacctgc agtatctctt cattgagaac aatgacatct gggcactatc caagttcacc 420
ttccgaggac tcaagtcctt gactcacctc tcgctggcca acaataacct gcagacactg 480
cccagagaca tcttccggcc cctggacatc ctgaatgact tggacctgcg gggcaactca 540
ctcaactgtg actgcaaggt gaagtggttg gtggagtggc tggcacacac caacaccacg 600
gtggcaccca tctactgcgc cagcccgccc cgcttccagg agcacaaggt gcaggacctg 660
ccgctgcggg agttcgattg catcaccaca gattttgtgt tgtaccagac cctggccttc 720
ccagcagtgt cggctgagcc cttcctctac tccagtgacc tctatttggc tctggcccag 780
ccaggagtca gtgcctgcac catcctgaag tgggactatg ttgagcggca gcttcgagac 840
tatgatagaa tcccagcccc ctctgcagtg cactgcaagc cgatggtggt ggacagccag 900
ctgtacgtgg tcgtggccca gctgtttggc ggctcttaca tttaccactg ggatcccaac 960
accacgcgct tcaccaggct gcaagacatt gacccgcagc gcgtgcgcaa gcctaacgac 1020
ctagaagcct tccgatccga cggtgactgg tactttgccg tggctgacag ctccaaggca 1080
ggcgccacca gcctctaccg ctggcaccag aatggcttct actcccacca ggcactgcac 1140
ccctggcacc gtgacaccga cctggagttt gtggacggcg agggcaagcc acggctgatt 1200
gtgtccagca gctcccaggc acccgtcatc tatcagtgga gtcgcacccca gaagcagttt 1260
gtggcccagg gtgaggtgac ccaggtgcct gatgcccaag ctgtgaaaca ctttcgtgcc 1320
ggccgcgaca gctacctgtg cctcagccgc tacattggcg actccaagat cctgcgctgg 1380
gagggtaccc gcttctcgga ggtgcaggcc ctgccctccc ggggctcgct ggccctgcag 1440
cccttccttg tgggtggccg ccgctacctg gcactgggca gtgatttctc cttcacccag 1500
atctaccagt gggatgaggg acgacagaag tttgtacggt tccaggagct ggctgtgcag 1560
gctcctcggg ccttctgcta catgcctgct ggggacgccc agctactcct ggcccccagc 1620
ttcaagggac agacgctggt gtatagacac attgtggtgg atctcagtgc ctagggggtg 1680
ccgaggcctc tggtctcctc agggtggcca ctggaggatg ggtgggggcc tctgtgagca 1740
ccacgtatgc ctgtgtgaag acaagcgacc aacctgcatt catgcccttg tatccgcaca 1800
cgcccagtgt gtacaggcgc accccgtgga ctggcccagg aggacgtgca tcatcactgc 1860
aatcagcatg aacaccaggc acccctggga cacctaacgc ccagttcccg tgaccccaac 1920
tgcaggcccc ccctactctg cagcctttcc catgctctgc ccctgtctc atgcacgagt 1980
ggtgcggggt gctgggagcg aggggggccg catcctcctg tccttgtgct tctccttctg 2040
cccctctggt ctttcctgtt ggtgctccag gcattgttta cctgctcttg ctccccgctg 2100
tagcccacag ccatatctgt catgcttccc cggggccact cacccctgcc ccaccccgtc 2160
```

```
tgcatgctca catggacgcg gacggacaca ctcattacac actcacacgc cccgactgca 2220
cccggtgtcc atctataaac atgtcagggc acgtgtgcat acaagcagat tcagcactgc 2280
ctaccaggct ctcctgttac cttgcctctc tgcttggagg ggccccctc tgctcacccc 2340
cagcgtgccc cccccggaac tgataaggat gagaatggtg gtgccagctt ctgagggcct 2400
gcttggcctc ctgggggcaa agccctctg cccgaagcaa taacaacagc agcagaagca 2460
atcccgggga gctgccaggg gtgtctgtct ctgttctcct tggcgtttcc ttcatggccg 2520
tgaatcccag ctcacccatg gacagaagac ctccctcat ctgaaagcct ttctgccttc 2580
cttcctgtag ctgtcgcttt ctcaggcttg gactgtcccc tggttcaccc tgtctttggg 2640
atgtggccct tgagctgggc gggctccttg ggtggattag agagctggcc gagctgcaag 2700
cacctgagtt acctgcgagg gtgcccagag tggagcttgg cctgagggcg tgttgcagcc 2760
ctggccccac cagtccacag aggcccccag gagcactgcc tttcagtgga ccgggtgggg 2820
gctggtggga ggggcccccac tagctgagct tctgctgccc tctgctgtct gctgagatgc 2880
caggctagtg gagagggcac accatgtgcc caaatggtgt gatgctcctt gtagcaggat 2940
accagggtac tgggggggcaa tgctatgatt aacttgcttc aaataaaaag ttcccgcccc 3000
ataaaaaaaa aaaaataacg catccacagc aaccacgcac caccatccac agccacttcc 3060
ccaccaggat tctctcacca accgctgtcc ccggcctatc acacacccag tcataccgtc 3120
gtgtatacct ctgtcccact cagctacaga catcggcgcc cccctcccc tccaccaccc 3180
tcttccatct                                                         3190
```

<210> 30
<211> 3150
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1629617CB1

<400> 30
```
ggcgcaccct acgacgcgca caccaccggc atgaccggcg ccatcagcta ccacccgtac 60
ggcagcgcgg cctacccgta ccagctcaac gaccccgcgt accgcaagaa cgccacgcgg 120
gacgccacgg ccactctcaa ggcctggctc aacgagcacc gcaagaaccc ctaccccacc 180
aagggcgaga agatcatgct agccatcatc accaagatga ccctcaccca ggtctccacc 240
tggttcgcca acgcgcgccg gcgcctcaag aaggagaaca agatgacctg ggccccgaga 300
aacaaaagcg aagatgagga cgaggacgag ggcgacgcta ccagaagcaa ggacgagagt 360
cccgacaagg cgcaggaggg cacggagacc tcggcagagg acgaagggat cagcctgcac 420
gtggactcgc tcacggatca ctcgtgctcg gccgagtcgg acggggagaa gcttccgtgc 480
cgcgccgggg acccctgtg cgaatcgggc tcggagtgca aggacaagta tgacgacctg 540
gaggacgacg aggacgacga cgaggagggc gagcgggggcc tggcgccgcc caagcccgtg 600
acctcgtcgc cgcttaccgg cttggaggcg ccgctgctga gcccccgcc cgaggccgcg 660
ccccgcggtg gccgcaagac gccccagggc agccggacgt ctccgggcgc gccgcccccc 720
gccagcaagc ccaagctgtg gtcgctggcc gagatcgcca cgtcggacct caagcagccg 780
agcctgggcc cgggctgcgg gccaccgggg ctgcccgcgg ccgccgcgcc ggcctcaacc 840
ggggcaccgc caggaggctc gccctaccct gcctcgccgc tgctgggccg cccctctac 900
tacacgtcgc ccttctacgg caactacaca aactacggga acttgaacgc ggcgctgcag 960
ggccagggtc tcctgcggta caactctgcg gccgcgagcc ccggcgaggc cctgcacacc 1020
gcgccaaagg cggccagcga cgcgggcaag gcgggcgcgc acccgctcga gtcccactac 1080
cggtccccgg gcggcggcta cgagcccaag aaagatgcca gcgagggctg caccgtggtt 1140
ggcggggggcg tccagcccta cctatagaag ggccgagcac agcaatgcaa gtaagtgggc 1200
accctggtgc ttcccagacc ctgtggagcc tcccagcacc aacttgctta atgtgttttt 1260
cttgcttctg gtcatggggt tagaagacag agcatttcaa tcttgtccag ctttttctttc 1320
cttaaggggg aaactatatt ttctgtttta tagattccct ctttttggca caagttctat 1380
cactccttta cctagggaag gaaaaaaaaa tatatatata tatgtgtgtg tgtgtgtgtg 1440
taatatatat atgtatatat atagtacatg tatatatata gtacatgtgt atatatatat 1500
tacacactac tgttgttttc ttttgaaaca atagattatt aaaatctggg catttatatg 1560
caatttagag ccattttac tatgctagac acataataga cacatgcttt gatgagcttg 1620
ggctaccaaa acttggagta gcaaaaacct ttatagatga tgtttgcaaa ataagactga 1680
agagacacga ggcgtactag tcttttatca gccctgtcca acctcccctc cacccatgtc 1740
caagtggaca ggactctgac attctcccgg ggctcacaac agtatctcaa aaaatatgtt 1800
gctttgcaac gagtcagaat tactaatcat agatcttgaa gataagaaag atttggaacc 1860
acttcagatc aactcgtttt tataatggtt ttgttgtgga actgaaagat tggggatgat 1920
ttggtttttt ccttggtcca caggtaggtg tcacaattgc tttgaaaaaa aaaaaaaaaa 1980
aaaaagtca gcaggccatt ctctcttccc aagctcataa atttacagat acaacaaaca 2040
```

```
gatgtctaaa tccagaccac atcttgtgcc actgtaaaag agaaggatcc acacagcact 2100
gtcagcccac agactctaac aattggaaca gactattgtt ggacatatgt gaatttgttg 2160
gcatagtata gtttccccga tgtatgtgtg acttttgaaa acaatctgtt tttctcagga 2220
tatcttgaat tgatcacttc attgccacgg tatatattct ataggtgtga taggatttac 2280
tgtaaagttt atttgtaaaa gatgtacaca gtagaaataa aacgtgatgg aagaacttga 2340
gtacttcaga agtggaaaca aatttgatat ttattttat aatgatataa agcttctagt 2400
aatttatgca agttgtattg caatggaatc taaacttttt gtaaataaat tctgcctggt 2460
ttttatttga acattgctgg ttatacaatc tttgttggtt gtttaacacg aattctttac 2520
taatttatat cttaaattgt aaataaaaac agactagtct acaataaaaa aaaaaaaaa 2580
gggcggccgc cgactagtga gctcgtcgac ccgggaatta attccggacc ggtacctgca 2640
ggcgtaccag cttttccctat agtgagtcgt attagagctt ggcgtaatca tggtcatagc 2700
tgtttcctgt gtgaaattgt tatccgctca caattccaca caacatacgg aagcatggtt 2760
tggactgcgc caagcgagaa ccaaggaatg gccaccgcgg aaccgcgggg cgggtcgagg 2820
gcccaaggca cacgagaa ggcccgcagg cggagcacta acacagtacc gcatactaaa 2880
gcgcgtgaac acgacttacg ctagacatac agggcggatg cagtgagaca atatatgacg 2940
cacagcagtc ggaagcacga ggcagatacc gtcacggagg caagtgcaca atacaacgac 3000
actaacaccg caacgacgag gcgccagaag cagcagtgac tatcaccgca caggccacac 3060
gtaaaatgca caaccacccg ccacgaccca cacaagaatc caaacgcgta ccgccgcgac 3120
cacagacaca gaacaccacg cacaagccaa                                  3150
```

```
<210> 31
<211> 2012
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2369279CB1

<400> 31
aaggggggagg ttactggtgg tggcggcggc ggcggcggcc cgggctccgg ctgtcgggga 60
agctgctgag ctgtgatggt gatgacgagg tgaaaatggc ggatctttcg aaatacaatc 120
ccggcccctg acataccaga ggcggcggcg acggcgacgt cgccaccccg gctcccttct 180
cggtccccgg gtaccctgga gcgctccagt ttggacaaac tgggaaagga agctgctggc 240
cagagctatc gtgcacgggc aaaacaagcc ccgccggggc tcgggattgc ccaggacctt 300
ctggagcccc tacctcggag cccgagggag gagaaagcgg cagccgctgg agtgcctgat 360
caactttcta gcgggagaac gatcatggag gtggtgccag ctgagattgc taagaagcct 420
cgaacgccaa cctctggtcc agtaatcacg aagctgatct ttgcaaaacc aattaatagt 480
aaagcagtta caggacagac aactcaagtt tcaccaccag tcattgcagg tagggttctt 540
tcacagtcta ctcccggaac tccatcaaag accataacaa tatctgaaag tggtgttatt 600
ggatcaactt taaattctac aacacagaca ccaaataaaa tagccatctc acctttgaaa 660
tcgccaaata aggcagtgaa atcaactgtg cagaccatca ctgttggagg agtgagcaca 720
tcacagttta agacaattat tcctctggca actgctccca atgtccagca gattcaagtg 780
cctggaagca agtttcatta tgtccgactt gttactgcca catcagccag tagctcaacc 840
cagccagtta gtcagaatcc cagtacaaac actcagcctc ttcagcaagc aaagccagtg 900
gttgttaata caaccccagt gcggatgtca gttccaattg tctcagctca ggctgtcaaa 960
caagttgttc caaaaccaat caatccaact tcacaaatag taactactag ccagccacag 1020
caacggctta tcatgcctgc cacaccactg ccacagatcc agcccaacct cactaacctg 1080
ccaccaggca ctgtcctggc accagctccg ggaacaggga atgtgggtta tgcagtgctt 1140
ccagctcagt atgttactca gctacagcag tcttcatatg tatcaatagc aagcaactct 1200
acctttactg gaacatctgg tatccagacc caggcacggc ttccattcaa tggcataatc 1260
ccatcagagt cggccagtcg gccccgaaag ccctgtaatt gtacaaaatc actgtgtttg 1320
aaattgtatt gtgattgctt tgcaaatggt gaattttgca acaactgcaa ttgtactaat 1380
tgttacaaca atttggaaca tgaaaatgaa aggcaaaaag caataaaggc atgccttgac 1440
agaaatccag aagcctttaa gcctaagata gggaaggaa aggagggaga atctgatcga 1500
cgtcatagca aagggtgtaa ttgcaaacga tcaggatgtc ttaaaaacta ctgtgaatgc 1560
tatgaggcaa aaataatgtg ttcctcaata tgcaaatgta ttggctgtaa gaattttgaa 1620
gaaagcccgg aaaggaagac attgatgcat ttggcagatg cagctgaagt aagggtacag 1680
caacaaacag cagccaagac gaagttatcc tctcaaattt cagacttgct tactaggcca 1740
acaccagctt taaatagtgg aggcggaaaa ttgccattta catttgtaac taaggaagta 1800
gctgaagcca catgtaattg cctccttgcc caggcagagc aggcagacaa gaagggaaaa 1860
tcaaaggcag cagcggaacg gatgatactt gaggaattcg gacgatgttt gatgagtgtc 1920
atcaactctg caggaaaggc aaaaagtgac ccttgtgcca tgaattgcta actcttgcac 1980
```

aaaagactga taaaatggaa ctgtacagaa aa   2012

<210> 32
<211> 1606
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3167506CB1

<400> 32
aggcggtgca atacacgggt tggctggcga tagcggcggt gggcgtggct taggcgcgag 60
cgagagcggg cgcggcgctc aggcagcggc ggttggctgc ggcgccagtg gtagtgctcc 120
aggctcgacg gcaccctcac agcgcccgcc cggccctgcc gctcatggcc gccccgccgc 180
gccccgcgcc atcgcccccc gccccgcggc gcctcgacac gagcgacgtc ctgcagcaga 240
tcatggccat caccgaccag agcctggacg aggcacaggc cagaaagcat gctctgaatt 300
gccatcggat gaagcctgct ctgttcagcg tgctctgtga gatcaaggaa aagacagtgg 360
taagcatccg tggcattcaa gacgaagatc cccctgacgc ccagctcctg aggctggata 420
acatgctgct ggctgagggc gtgtgcaggc ccgagaagag aggaagagga ggagcggtgg 480
ccagggccgg cacagcaaca ccaggtggct gtccaaatga caatagcatt gagcactctg 540
actacagggc caagctgtcc cagatccgac agatttacca ctctgagcta gagaaatatg 600
aacaggcctg tcgtgagttc accacgcacg tcaccaacct cctccaggag cagagcagga 660
tgaggcctgt ctcccctaag gagattgagc gcatggtcgg cgccattcac ggcaagttca 720
gcgccatcca gatgcagttg aagcagagca cctgtgaggc agtgatgacc ctgcgttcgc 780
ggctgctcga tgccaggcgc aagcggcgga atttcagcaa gcaggcgacg gaagtgctga 840
atgagtattt ttactcccat ctgaacaacc cttaccccag cgaagaagcc aaagaagagc 900
tggccaggaa gggcggcctc accatctccc aggtctctaa ctggtttggc aacaaaagaa 960
tccggtataa aaagaacatg gggaagtttc aagaagaggc taccatttac acgggtaaaa 1020
cggctgtgga taccacggaa gttggggtcc cagggaacca cgccagctgc ctgtcaacac 1080
ctagctccgg ctcctctgga cccttccgc tgcccagcgc tggggacgcc ttcctcaccc 1140
tgcggactct ggcctctctc cagcctcctc ctggggggagg ctgcctgcag tcccaggccc 1200
agggtagctg gcagggggcc accccccaac ctgcaactgc ctcacctgct ggagaccctg 1260
gcagcatcaa ctccagtaca tctaattaag tttgggggat aagcaggaaa gagcgctgcg 1320
tgagctgcca tgtatcgcca gccgttgctt tgttactgaa cgtgccgccg acgacctcag 1380
aaaacccaga tgggtggtgg tgcccatgag cccctgctcc tcagccaggc ccgtggcgcc 1440
ggctcatgtg tctgctgcga ctcgagatgg cctgaaacgc cactcattct cccacttcag 1500
ttcgtttttt tgacagtaat tttatggtaa cgctatgaat tgaattgtct gttctaggac 1560
tgggcacaga ttttcccatt aaaatttttg acttatttta aaaaaa   1606

<210> 33
<211> 7364
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3075937CB1

<400> 33
cccctaacgg agcccccagt gtgtgtgccg aagagggctc cctggggccc cgcaacgccc 60
gcagccagcc ccgcaagggc gcgtctgatt tgccaggtga gcctccggcc gagggcgcag 120
cgcacacggc cagctccgcg caggcagact gcacagcccg ccccaagggt cacgcccacc 180
ctgctaaggt gctaactttg gacatctact tgagtaagac tgaggggggca caagtggacg 240
agccggtcgt gattactccc agagcggaag attgcggtga ctgggacgac atggagaaga 300
ggtccagcgg ccgtaggtcg gggaggcgga ggggtcgca gaaatccacc gactccccg 360
gcgcggacgc cgagctccct gagagcgctg ccagggacga cgcggtgttc gacgacgagg 420
tggcgccaaa cgcggccagc gataacgcct cggcggaaaa gaaagtgaaa tctccgcggg 480
cagccctcga cgggggcgtt gcctccgctg cgagcccaga gtccaagccc agccccggta 540
ccaaagggca gctccgaggg gagtcggacc ggagcaaaca gccacccccg gcttcgtccc 600
ccacgaagag gaagggcagg agccgtgccc tcgaggccgt gcccgccccg cccgccagcg 660
gcccccgggc tcccgccaag gagtccccac ccaagagggt gcccgatccc agcccagtca 720
ccaagggcac tgcggccgag agcggggagg aggcggcgcg ggccatctcc cgcgagctcc 780

```
cggtcaagag cagctcgctg ctgccggaga tcaagcccga gcacaagagg ggcccgctcc 840
ccaaccactt caacggccgg gcagagggag gtcgaagcag agagctgggc agagcggccg 900
gagcgcctgg agcttctgac gccgacggct tgaagcccag gaaccatttc ggcgtgggca 960
ggtcgacagt gaccactaaa gtgaccctcc ctgccaagcc caaacatgtg gaactaaatc 1020
ttaaaacccc taagaatctt gacagtttgg gaaatgagca caatccattt agccagccag 1080
ttcacaaagg caacactgcc accaaaatct ccttatttga aaacaaacgg acaaacagta 1140
gcccaagaca cactgacatt cgaggcccaa ggaatactcc tgcctctagt aaaacgtttg 1200
ttgggagggc aaagctgaat ttagccaaaa aagccaaaga aatggagcaa cctgaaaaga 1260
aagtaatgcc aaacagtccc cagaatggtg tgctggttaa ggaaactgct atagaaacca 1320
aagttaccgt ctcggaagaa gagattctgc cagcaaccag aggaatgaat ggagactctt 1380
ctgagaatca agctcttggt cctcagccta accaagatga taaagcagat gtacaaacag 1440
atgctggctg cctttcagaa ccagtggctt ctgctctgat tcctgtcaag gatcataagc 1500
tcttagagaa ggaggactca gaggctgcag acagcaaaag ccttgtactt gaaaatgtaa 1560
ccgatacagc acaagacatc cccaccactg tggataccaa agatttacct ccaacggcca 1620
tgccaaagcc acagcataca tttttctgact cacagtcccc tgctgagtca tctcctgggc 1680
cttctctttc actgtctgca cccgctcctg gggatgttcc caaagacaca tgtgttcaat 1740
cacccataag cagtttccca tgcactgatc taaaagtgtc agaaaaccat aaaggatgtg 1800
ttttgcctgt gtctcgtcag aacaatgaga aaatgccact tttagaactt ggaggagaaa 1860
caaccccctcc tttgtccaca gagcgtagtc cagaagctgt gggaagtgag tgtccatcca 1920
gagtcctcgt ccaggtcagg tccttcgtgc tccccgtgga gagcacccag gatgtgagct 1980
cccaggtcat cccagagagc tctgaagtta gagaagtgca gttgccaact tgtcacagta 2040
atgaacctga agtggtttcc gttgcaagtt gtgctccccc acaagaggaa gtactgggca 2100
atgaacactc tcattgcaca gcagagctcg cggcaaaatc tggcccacaa gtcataccgc 2160
cagcatcaga gaaaactctg cctattcagg ctcaaagtca gggcagcaga acacccctga 2220
tggctgaatc cagtcccacc aactctccca gcagcggaaa tcacttagcc actcctcaaa 2280
ggccagatca gactgttaca aatggccagg atagccctgc cagccttttg aacatttctg 2340
ctggtagtga tgatagtgta tttgattctt cttctgatat ggaaaaattc actgaaatta 2400
taaaacagat ggatagcgca gtttgtatgc ccatgaaaag aaagaaggcc aggatgccaa 2460.
actctcctgc tcctcacttt gccatgcctc ctattcacga agaccattta gaaaaggtgt 2520
ttgatcccaa agtgtttacc tttggtttgg ggaagaagaa ggaagtcag ccagaaatgt 2580
caccggcttt acatttgatg cagaaccttg acacaaaatc caaactgaga cccaaacgtg 2640
catctgctga acagagcgtc ctcttcaagt ccctgcacac caacactaat gggaacagtg 2700.
agcctctggt gatgccggaa atcaatgaca aagagaacag ggacgtcaca aatggtggca 2760
ttaagagatc gagactagaa aaaagtgcac ttttctcaag cttgttatct tctttaccac 2820
aagacaaaat ctttttctcct tctgtgacat cagtcaacac tatgaccacg gctttcagta 2880
cttctcagaa cggttcccta tctcagtctt cagtgtcaca gcccacgact gagggtgccc 2940
cgccctgtgg tttgaacaaa gaacagtcaa atcttctgcc cgacaactcc ttaaaggtct 3000
tcaatttcaa ctcgtcaagt acatcacact ccagtttgaa aagtccaagc cacatggaaa 3060
aatacccgca aaaagagaaa accaaagaag atctggattc acgaagcaac ctacacttgc 3120
cagaaactaa attttctgaa ttgtcaaaac tgaagaatga tgatatggaa aaggctaatc 3180
atattgaaag tgttattaaa tcaaacttgc caaactgtgc aaacagtgac accgacttca 3240
tgggtctttt caaatcaagc cggtatgacc caagcatttc tttttctgga atgtcattat 3300
cagacacaat gacacttaga ggaagtgtcc aaaataaact caatccccga cctggaaagg 3360
tagtgatata tagtgaaccc gacgtctctg agaagtgcat tgaagttttc agtgacattc 3420
aggattgcag ttcttggagc ctctctccag tgatactcat aaaagttgtt agaggatgtt 3480
ggattttgta tgagcaacca aattttgaag ggcactccat ccccttagaa gaaggagaat 3540
tggaactctc tggtctctgg ggtatagaag acattttgga aaggcacgaa gaagcagagt 3600
ctgataagcc agtggtgatt ggttccatca gacatgtggt tcaggattac agagttagtc 3660
acattgactt atttactgaa ccagaagggt taggaatcct aagttcctac tttgatgata 3720
ctgaagaaat gcagggattt ggtgtaatgc agaagacttg ttccatgaaa gtacattggg 3780
gcacgtggct gatttatgaa gaacctggat ttcagggtgt tcctttcatc ctggaacctg 3840
gtgaataccc tgacttgtcc ttctgggata cagaagcagc gtacattgga tccatgcggc 3900
ctctgaaaat gggtggccgt aaagttgaat ccctacaga tccaaaggta gttgtttatg 3960
aaaagccttt ctttgaagga aaatgtgtgg aactagaaac aggaatgtgt agttttgtca 4020
tggagggagg tgaaacagaa gaggcgactg gagacgatca tttgccgttt acgtcagtgg 4080
ggtctatgaa agttctaaga ggcatttggg ttgcatatga gaagcctgga tttaccggtc 4140
atcagtattt gctagaagaa ggagaataca gggactggaa agcctgggga ggttacaatg 4200
gagagcttca gtctttacga cctatattag gtgattttc aaatgctcac atgataatgt 4260
acagtgaaaa aaactttgga tccaaaggtt ccagtattga tgtattggga attgttgcta 4320
atttaaagga gactggatat ggagtgaaga cacagtctat taatgtactg agtggagtat 4380
gggtagccta tgaaaatcct gacttcacag gagaacagta tatactggat aaaggatttt 4440
ataccagttt tgaggactgg ggaggcaaaa attgtaagat ctcttctgtt caacctatat 4500
gtttggattc tttcactggc ccaaggagac gaaatcagat tcacttgttt tcagaaccac 4560
```

```
agtttcaagg tcacagtcaa agttttgaag aaacaacaag tcaaattgat gattcatttt 4620
ctaccaagtc ttgcagagtt tcaggaggca gctgggttgt atatgatgga gaaaatttca 4680
ctggtaatca atacgtgttg gaagaaggcc attatccttg tctgtctgca atgggatgcc 4740
cgcctggagc aactttcaag tctcttcgtt ttatagatgt tgaattttct gaaccaacaa 4800
ttattctctt tgaaagagaa gacttcaaag gaaaaaagat tgaacttaat gcagaaactg 4860
tcaatctccg atccctggga ttcaacacac aaatacgctc tgttcaggtt attggtggca 4920
tatgggttac ttatgaatat ggcagttaca gagggcgaca gttcctattg tcacctgcag 4980
aagtacctaa ttggtatgaa ttcagtggct gtcgccaaat aggttctcta cgaccttttg 5040
ttcagaagcg aatttatttc agacttcgaa acaaagcaac agggttattc atgtcaacca 5100
atggaaactt agaggatctg aagcttctga ggatacaggt catggaggat gtcggggccg 5160
atgatcagat ttggatctat caagaaggat gtatcaaatg caggatagca gaagactgct 5220
gcctgacgat tgtgggcagc ctggtaacat ctggctccaa gctaggcctg gccctggacc 5280
agaatgctga cagccagttc tggagcttga agtccgatgg caggatttac agcaagttga 5340
agccaaattt agttttagac attaaagggg gcacacagta tgatcaaaat cacattatcc 5400
tcaacactgt cagcaaagag aagtttacac aagtgtggga agccatggtc ctatatacct 5460
gaacaaagaa ggaagaagaa tcttctggag gtccttccag ccaccttatt tcttaaaaag 5520
gacaatgctg atggaagacc agactggaaa gtggatcgac tcctccttca ttgattctaa 5580
attcaacctt aaatcatgct gccatgactc agagaactta ctcatcgttt caaaagacta 5640
tcatagcttt aaaccaataa tttgtcctcc tttcatttct tgcctttcat ttttggtagc 5700
tgcttaaaca ggttgcctaa ttagcagctt ttgggtgatt ttgtaaaatg ttatatcaag 5760
atttcaagac tgtgtacatt ttaaattatt ccaaagata gtgacaggag agaactggaa 5820
caaatttacc aactttgtgg acctacaaag cccttacact ttaaagggta agacaaaggc 5880
ttaagtttga aaggtagaga actgtttagc atctgagaag aaatacttta ttaggcctgt 5940
aattttggtt cttggcctta aacactttct ggaaccttta aatatgctgc atagcacaat 6000
gggaaagcct taggtattca cacatttaag gaactctaaa caaaatacta ttttcctttа 6060
gttcatatta aaaattaata cattttaaaa atttaatgtc aaagtctggt aacatttgtt 6120
agtaggattt gagttattat ttttgagac aggatctcag gctggagtgc agtggcacaa 6180
tcacggctca ctgcagcctc tacctcccca ggctcaggtg atcctcccac ctcagcctcc 6240
caagtagctg ggactatagg cacacatcac caagcccagc caaattttgt ttttttttg 6300
tagagatggg gtttcatcac gttgcccagg ctgatctcga acctctggc tcaagcaatt 6360
cactcgcctc ggcctcccaa aatgctggga ttacaggcct gagccactgc gcccagccag 6420
gatttgaatt attttaactc atccatgggc tgccctagaa tgtcacaaat gagggttgtt 6480
taatgccttt cttatagctg ctactggaac actattatga cctaatttat gagccatcct 6540
tactcatcta caagtgctga agcaatgtta catacttttt tgctaaactc agattttta 6600
gcctaatttc ttgtcctcct atccacctgc atccacacat ggcctgcatg gggctgcctt 6660
ccctgcagtg ttctgcagcc atgcttcagg gtatagctgt tggtggacag cctcaggtct 6720
tgggggcact atagccacta aacgaggtgt gaaaggctca agaggatgac cagcaattaa 6780
ttatccccag aaagtgaagg aaaagagacc tttagggatg ttgctggtca agtcttgatt 6840
tgaccggagt caaatcaatc ttcaagcaat cttggaatcc tcaactgcag taagcatttc 6900
aaaatgcaaa caaactgctt aacaactgac aagacaccag cccatacgct gctcttccaa 6960
cagtgggttc tagctttgaa caaaagtgct aaacatttcc ttgaatatat tcttcctctt 7020
tttgtcctca tcactcaata ctggtgctct tgtcacaggt agaacagctt gtttctttc 7080
catctattca agtgtgtttc taattctaaa atgctgatct tctctggagt ctatggtagg 7140
caattatggt cactggaata gtttgtcttg ttttaaaata ttattggtgc atgtacaaca 7200
gcatccaaca tatctgtctt gttcctagat atatagctct gatttaggc cttttgtgca 7260
taccattaca atatggtggg gtaagacatt ctacagtagc ctgtgctgaa ctgatctctt 7320
aaataaactt gcttctggtt aactatgaaa aaaaaaaaa aaaa              7364
```

```
<210> 34
<211> 1345
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5176268CB1

<400> 34
gtgggattaa gaagtgatag gagagcttgc tgtctgtctc tgctctccac tgtgtgagga 60
tacaacagga agacagccat ctggtgagga agagagggcc ctcgccagat accggacctg 120
ctgacacctt gatcttggac ttcccatctt ccaggaaggc ctgacctcag ttgttccagg 180
gtaaagaatt tgggcagtgc ccacacccac gctgttggat aacatttctt caccatacca 240
gtgagggtga atgtgtacac gcccagcttc ctgcctgtta ctctccacag tatgcgaaga 300
```

```
atatccctga cttctagccc tgtgcgcctt cttttgtttc tgctgttgct actaatagcc 360
ttggagatca tggttggtgg tcactctctt tgcttcaact tcactataaa atcattgtcc 420
agacctggac agccctggtg tgaagcgcag gtcttcttga ataaaaatct tttccttcag 480
tacaacagtg acaacaacat ggtcaaacct ctgggcctcc tggggaagaa ggtaaatgcc 540
accagcactt ggggagaatt gacccaaacg ctgggagaag tggggcgaga cctcaggatg 600
ctcctttgtg acatcaaacc ccagataaag accagtgatc cttccactct gcaagtcgag 660
atgttttgtc aacgtgaagc agaacggtgc actggtgcat cctggcagtt cgccaccaat 720
ggagagaaat ccctcctctt tgacgcaatg aacatgacct ggacagtaat taatcatgaa 780
gccagtaaga tcaaggagac atggaagaaa gacagagggc tggaaaagta tttcaggaag 840
ctctcaaagg gagactgcga tcactggctc agggaattct tagggcactg ggaggcaatg 900
ccagaaccga cagtgtcacc agtaaatgct tcagatatcc actggtcttc ttctagtcta 960
ccagatagat ggatcatcct gggggcattc atcctgttag ttttaatggg aattgttctc 1020
atctgtgtct ggtggcaaaa tggcagaaga tccacctaga ggtgatacca cggcggcgca 1080
gagttgttca cctgtggtcc tcgatcgctg acagccttgg ctcccactgc tgtgtgttcc 1140
ctgagtcaag tggaggcgga gcctttgctc gtgagattgt ggacctatga tcctaccacc 1200
agcccacctg gttctaacac ccccctcctct atgtgtgaga gggagagaag aaaagtgagg 1260
gagaaaagag agataagcaa gaccagagag gaaaaatctg tccagcacct gcgccgtaac 1320
tagtgatcca gctcgtcccc acagc                                        1345
```

```
<210> 35
<211> 4255
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 72830854CB1

<400> 35
agccggttcc cccccacctg ccaccgtacc tggctccagc ctcccaggtg ggggcccccg 60
ctcagctgaa gccctccag atgccacagg cgcccctgca gccgcttgct caagtccctc 120
cgcagatgcc cccgattcct gttgtgcccc ccatcacgcc cctggcgggg aatcgacggc 180
ctccctccgg ccctcccaga cctgccgacc gcgactgtgc ctcccatgcc accacctcag 240
tatttctctc cagccgtgat cttgccgagc ctcgctgccc cactccccccc tgcgtcccca 300
gccttgcctc tgcaggctgt gaagctgccc caccccccctg gggcgcccct ggccatgccc 360
tgccggacca ttgtgccaaa tgcaccggcc actatccccc tgctggccgt agccccaccg 420
ggcgtggctg ccctgtccat tcattctgcc gtggcccagc tcccaggcca acctgtgtac 480
ccagcggcct tcccacagat ggcgcctact gacgtccctc cttccccccca tcacacggtg 540
cagaatatga gggccacccc tccacagccg gcactgcctc cacaacccac actgcccccca 600
caacccgtgc tgcccccgca acccacgctg ccccctcaac ctgtgttgcc cccgcaaccc 660
acacggcccc ctcaacctgt gctgcccccg caacccatgc tgcccccaca acctgtgctg 720
cccccgcagc cggcactgcc tgtgcgccct gagcccctcc agccccacct tcctgaacaa 780
gctgctccag ctgctacacc agggagccag attctgcttg gccacccagc tccctatgct 840
gtggacgtcg ccgctcaggt cccccaccgtg cctgtgccac cggctgcggt cctctcgccg 900
cctctgccgg aagtgctgct gcctgcccac cctgagctcc tgcctcagtt ccccagctcc 960
ctggccacgg tgtctgcctc tgtgcagagt gtgcccaccc agactgccac acttctgcca 1020
ccagcaaacc caccgctgcc tggcgggccc gggatcgcca gcccttgccc aactgtccag 1080
ctgacggtgg aaccagtcca agaggagcag gcctcacagg acaagccgcc cggcctcccg 1140
cagagctgtg agagctatgg aggttctgat gtcacttctg gaaaagagct gagtgacagc 1200
tgtgaaggcg cctttggagg gggcaggctg gagggcaggg cagcccgaaa acaccaccgc 1260
aggtccacgc gtgcgcgctc ccggcaggag agggccagcc ggccccggct taccatcttg 1320
aacgtgtgca acactgggga caagatggtg gagtgccagc tggagacgca caaccacaag 1380
atggtgacct tcaagttcga cttggacggg gacgcacccg atgaaattgc cacgtatatg 1440
gtggagcatg actttatcct gcaggccgag cgggaaacgt tcatcgagca gatgaaggat 1500
gtcatggaca aggcagagga catgctcagc gaggacacag acgccgaccg tggctccgac 1560
ccagggacca gccgccaca cctcagcacc tgcggcctgg caccgggga ggagagccga 1620
caatcccaag ccaacgcccc cgtgtatcag cagaacgtcc tgcacaccgg gaagaggtgg 1680
ttcatcatct gtccggtggc tgagcacccc gcccccgagg cccctgaatc ttcgcccccca 1740
cttcctctaa gctccctgcc gccagaagcc agccaagatt cagcgcccta taaagaccag 1800
ctgtcctcga aggaacaacc cagctttcta gccagtcagc agctcctgag ccaggcgggc 1860
cccagcaacc ctcctggggc acccccagcc cctttggccc cctcctcccc tcctgtgact 1920
gctctgcccc aagatggagc agctccagcc accagcacca tgccagagcc agcgtcagga 1980
actgccagcc aggcaggggg tccagggaca cctcaggggc tgaccagtga gctcgagacg 2040
```

```
tctcagccac tagcggagac tcacgaggcc ccgcttgctg tgcagcccct cgtggtgggc 2100
ctagcacctt gcactccagc tccagaggct gcctcaacca gggacgccag tgccccaagg 2160
gagcccctgc cacctcctgc acctgagccc agcccccaca gcgggacccc acagcccgcc 2220
ttgggtcagc ctgctcccct gcttcctgcc gcagtggggg ccgtcagcct ggccacctcc 2280
cagctcccaa gcccacccct ggggcccacc gtcccccac agccaccctc ggccctggag 2340
tcggatgggg aagggccgcc ccccagggtg ggctttgtgg acagcaccat caagagcctg 2400
gacgagaagc tgcggactct gctctaccag gagcacgtgc ccacctcctc agcctcagct 2460
gggaccctg tggaggtggg cgacagagac ttcaccctgg agcccctgag aggggaccag 2520
ccccgctcag aggtctgcgg ggggggacctg ccctgcccc cagtgcctaa ggaggcggtc 2580
tcagggcgtg tccagctgcc ccagcccttg gtggagaagc agaactggc ccccactcga 2640
ggggccgtga tggagcaggg cacgtcctcg tcaatgacag cagagtcgtc tcccaggagt 2700
atgctaggct atgacagaga tggaaggcag gtggcctcag actcccatgt ggtccccagc 2760
gtcccccagg atgtacctgc ttttgtgaga cctgcacgtg tggagcccac agacagggat 2820
ggtggagaag ctggagaaag ctcggcagag cccccgccga gtgacatggg cacagtgggg 2880
ggccaggcta gccacccca gacactcggc gctcgagctt tggggtcccc tcggaaacgt 2940
ccagagcagc aggatgtcag ctcaccagcc aagactgtgg gccgtttctc ggtggtcagc 3000
actcaggacg agtggaccct ggcctccccc cacagcctga gatactctgc cccacccgac 3060
gtctacctgg acgaggcccc ctccagcccc gacgtgaagc tggcagtgcg gcgggcgcag 3120
acggcctcct ccatcgaggt cggcgtgggc gagcccgtgt ccagcgactc tggggacgag 3180
ggcctcgggg cgagacccc ggtgcagaag caggcgtccc tgcccgtgag tggcagcgtg 3240
gctggcgact tcgtgaagaa ggccaccgcc ttcctgcaga ggccttctcg ggccggctcg 3300
ctgggccccg agacacccag caggggtgggc atgaaggtcc ccacgatcag cgtgacctcc 3360
ttccattccc agtcgtccta catcagcagc gacaatgatt cggagctcga ggatgctgac 3420
ataaagaagg agctgcagag tctgcgggag aagcacctga aggagatctc ggagctgcag 3480
agccagcaga gcaggagat cgaagctctg taccgccgcc tgggcaagcc actgcccccc 3540
aacgtgggct cttccacac ggcacccccc actggccgcc ggagaaaaac cagcaagagc 3600
aagctgaagg caggcaagct gctaaatccc ctggtgcggc agctcaaggt cgtggcctcc 3660
agcacagggt cctccaccag cagcctggcc ccaggccctg agccaggccc ccagcccgcc 3720
ctgcacgtcc aggcgcaggt gaacaacagc aacaacaaga agggtacctt cacggacgac 3780
ctgcacaagc tggtggacga gtggacgagc aagacggtgg gggccgcgca gctgaagccc 3840
acgctcaacc agctgaagca gacccagaag ctgcaagaca tggaggccca ggcaggctgg 3900
gctgccctg gcgaggcgcg ggctatgacc gcacctcgag caggagtggg gatgccacgt 3960
ctgccccag cgcccggccc tctgtccacc acggtcattc ccggagccgc cccgaccctg 4020
tccgtgccca caccagatcc tgagagtgag aagcctgact gaccccgcct agacgccagg 4080
cccacttcac gccgtctaag tggagaagtg acggaccctc agggccagct gctcctcctg 4140
tccagttcac gctgtttgt aaccactttc taagcatttt ttattcacaa ttggaaacac 4200
aaatgtaatg caagaataaa aaatattttg gggccaaaaa aaaaaaaaa agggc     4255
```

<210> 36
<211> 4181
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3632052CB1

<400> 36
```
ttgttatttc atatgtgctt tattctccct ctcctatcgt cttatgttat tttatcagcg 60
tccctttgaa tttattcttc tttcttctct tctgatttct gttattttttg tgtggttgtg 120
tgtgattttg ttatgcgtgt ctcgtcttct tattgtctct tcatttttctt cttgtttcat 180
tatgtcgttt tttgtctctt tttctattct ctccttctttt tctattttttt ctgttccgtc 240
ttttgtcttg tgtctcctgc cttcgcaaag ggggatgtgc tgccaggcga ttaagttggg 300
taacgccagg gttttcccag tcacgacgtt gtaaaacgac ggccagtgaa ttctcagttt 360
ggataattgg tcaaacaatt ttttatccag tagtaaaata tgtaaccatt gtcccagtaa 420
agaaaaataa caaaagttgt aaaatgtata ttctcccttt tatattgcat ctgctgttac 480
ccagtgaagc ttacctagag caatgatctt tttcacgcat ttgctttatt cgaaaagagg 540
ctttttaaaat gtgcatgttt agaaacaaaa tttcttcatg gaaatcatat acattagaaa 600
atcacagtca gatgtttaat caatccaaaa tgtccactat ttcttatgtc attcgttagt 660
ctacatgttt ctaaacatat aaatgtgaat ttaatcaatt ccttttcatag ttttataatt 720
ctctggcagt tccttatgat agagtttata aaacagtcct gtgtaaactg ctggaagttc 780
ttccacagtc aggtcaattt tgtcaaaccc .ttctctgtac ccatacagca gcagcctagc 840
aactctgctg gtgatgggag ttgtattttc agtcttcgcc aggtcattga gatccatcca 900
```

```
ctcacatctt aagcattctt cctggcaaaa atttatggtg aatgaatatg gctttaggcg 960
gcagatgata tacatatctg acttcccaaa agctccagga tttgtgtgct gttgccgaat 1020
actcaggacg gacctgaatt ctgattttat accagtctct tcaaaaactt ctcgaaccgc 1080
tgtgtctcca atatcttctt caggctctga caggcctcct ggaaacttcc acatattttt 1140
caatttattt cgatcttgta caaccagtat ttttctagta ctttcatcaa atacagctcc 1200
tgcaactcct acttgatgtg aagcatatcc tggtaatctg ctgggccctt ctctcagcca 1260
cagagtcaac gttgatgaat ccgattctgc gtggtgaaag cggaagccca gggaagcagc 1320
aggggcaata aatcggcttt ggaggatggg aatgtgcagc catacagctg ttctaccttc 1380
tgatcgccat tgctgtactg cagcctgcaa gcccttctgg aaggcggcag cgtccaggcg 1440
gtccagcgca tcgagccgcg ccaggcgcac cgagatgccc ccgaatctgt ccagctcgcc 1500
ctgcagatcg cacgctccaa ctggcggatt ccgcacgtaa ccctgtgcgc ccgaggccca 1560
gcggtaaccc gccgaagcgg gggcggtacc ttgtgaccat gtcttccaag aagaatagaa 1620
agcggttgaa ccaaagcgcg gaaaatggtt cgtccttgcc ctctgctgct tcctcttgtg 1680
cggaggcacg ggctccttct gctggatcag acttcgcggc aacctccggg actctgacgg 1740
tgaccaactt attagaaaag ggtaaagaat tccgggtgta tacagcctgg cctatggcag 1800
gatttcctgg aggcaaggtc ggcctgagtg aaatggcaca gaaaaatgtg ggtgtgaggc 1860
ctggtgatgc catccaggtc cagcctcttg tgggtgctgt gctacaggct gaggaaatgg 1920
atgtggcact gagtgacaaa gatatggaaa ttaatgaaga agaactgact ggttgtatcc 1980
tgagaaaact agatggcaag attgttttac caggcaactt tctgtattgt acattctatg 2040
gacgaccgta caagctgcaa gtattgcgag tgaaaggggc agatggcatg atattgggag 2100
ggcctcagag tgactctgac actgatgccc aaagaatggc ctttgaacag tccagcatgg 2160
aaaccagtag cctggagtta tccttacagc taagccagtt agatctggag gatacccaga 2220
tcccaacatc aagaagtact ccttataaac caattgatga cagaattaca aataaagcca 2280
gtgatgtttt gctggatgtt acacagagcc ctggagatgg cagtggactt atgctagagg 2340
aagtcacagg tcttaaatgt aattttgaat ctgccagaga aggaaatgag caacttactg 2400
aagaagagag actgctaaag ttcagcatag gagcaaagtg caatactgat actttttatt 2460
ttatttcttc aacaacaaga gtcaatttta cagagattga taaaaattca aaagagcaag 2520
acaaccaatt caaagtaact tatgacatga taggaggatt aagtagccag ctgaaagcaa 2580
ttagagaaat aattgaattg cccctcaaac agcctgagct tttcaagagt tatggaattc 2640
ctgccctag aggattgtta ctttatggtc ctccatgtac tggaaaaaca atgatcgcca 2700
gggctgttgc taatgaagtt ggagcctatg tttctgtaat taatggtcct gaaattataa 2760
gcaaattcta tggtgagact gaagcaaagt tacgtcagat atttgctgaa gccactctac 2820
gacacccatc aattattttt attgatgagc tggatgcact ttgtccgaaa agagaggggg 2880
cccagaatga agtggaaaaa agagttgtgg cttcactctt aacactgatg gatggcattg 2940
gttcagaagt aagtgaagga caagtgttgg ttcttggggc cacaaatcgc cctcatgcct 3000
tggatgctgc tctccgaaga cctgggcgat ttgataaaga gattgagatt ggagttccca 3060
atgctcagga ccggctagat attctccaga aactgcttcg aagggtaccc catttgctca 3120
ctgaggctga gctgctgcag ctggcaaata gtgctcatgg atacgttgga gcagacttga 3180
aagtcttgtg taatgaagca ggtgagtgtg gtttgctatg gtgagtctct attgatgcac 3240
ttatctccag tttacttaca tacaaataat ttatatttta cagatttctt aatggaagta 3300
gctttgtttc taattataaa atgtgtaatt tttatttgaa aaattttta ttggaaaaac 3360
ctagatgatt cagaagacta taaaaaaaaa taaaaatcac tcaatcctac ctctaaagga 3420
taaccaccat tatagtgtat acagtcagcc ctccatatcc ttgggctcag catctgagga 3480
gtcaaccaac tgtggtctga aaatattcag gaaaaaaaag ttaaaacagt aaaaaataat 3540
aaaaacttaa aaaatatagt atcacaacta cagttgaccc ttgaactatg agggcttgaa 3600
ctgcatgggt ccacttatat atgaatattt ttcaataaat atattggaaa attttttgga 3660
gatttgcaac aatttgaaaa agcttgcttt ttctttagct tactttattg taagaataca 3720
gcatataatg catataaaaa atatggttaa tcagctgttt atattattgg caaggcttca 3780
ggtcaacaat aggctattag tagttagttt ttgggggagt cagaagttat gcatagtttt 3840
ttgactgcac agggattggt gtccctgacc cccaagttgt tgaggggtca actgtattta 3900
catagcattt acatgggatc aggtaaaagt aatctagaga attattttaa cttatacggg 3960
aggacgtgtg ttagccaatt ttgcacagta taatgccctt aagggtttag acgccttaaa 4020
tttgggatct taaggggccg ggacaaataa tcccagaaat aggggaaaat cgctaaaat 4080
ttccaagatt atacacaaaa gaacacccga agggttgttt ttttcccca catcttctct 4140
taaaattaat accggtccaa cctcggaaga aaaaattatt g                    4181
```

<210> 37
<211> 2478
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

```
<223> Incyte ID No: 7493817CB1

<400> 37
gggctgcaga aagaggaggc caggagcggt cccatccgtc ccgtcccgtc ccgtctcccc 60
ctcttcctct tgctccttgc cccccggctc tgcgagagtt gagggttcag gtggccgtac 120
gcggcagtga gggcaagagg gccggggagag tggggagcgg aggcaggagt gcggggggaag 180
atgcccatcc tgctgttcct catagacacg tccgcctcta tgaaccagcg cactgacctg 240
ggcacctctt atttggacat tgccaaaggc gctgtggagt tattcttgaa gctgcgcgcc 300
cgggacccgg ccagccgtgg agacaggtac atgctggtca cctacgacga accccccgtac 360
tgcatcaagg ctggttggaa ggaaaatcat gcaacattca tgagcgaact aaaaaaatctt 420
caggcttctg gactgactac tctcggtcag gctctaagat cctcatttga tttgttaaat 480
ctcaatagat taatatctgg aatagacaat tatggacagg ggagaaatcc attttttta 540
gaaccatcta ttttaattac catcacagat ggaaacaagt taacaagtac tgctggtgtt 600
caagaagagc tccatcctcc tttgaattcc cctctgcctg gaagggaact aaccaaagaa 660
ccttttcgtt gggatcaaag gttatttgcc ctggtgttgc gtttgcctgg agtggcttct 720
accgaaccag agcaactagg gagcgtacca actgatgaat ctgccatcac acagatgtgt 780
gaagtcacag gaggtcgctc ctactgtgtg agaacacaaa gaatgttgaa tcaatgttta 840
gaatctctag ttcaaaaagt tcagagtggt gtagttatta attttgaaaa aacaggacca 900
gatccacttc ctattggaga agatggactt atggattcat ccaggccaag caattcattt 960
gctgctcagc catggcatag ttgtcataaa ctcatttatg tacgacctaa ctctaaaact 1020
ggtgttcctg ttggacattg gccaattcca gaatcttttt ggccagatca gaatttacct 1080
tcactacctc cacgaacatc tcatcctgtt gtgaggttct cctgtgtaga ttgtgagcca 1140
atggtaatag acaaacttcc ttttgacaaa tatgaacttg aaccttcgcc cttaactcag 1200
tatatcttgg aacgaaagtc tccccatacc tgctggcagg tatttgttac tagcagtgga 1260
aagtacaatg aacttggata tccatttggc tatttaaaag ccagtacaac tttaacttgt 1320
gtaaacctct ttgtgatgcc ttacaactac ccagttttac ttcctctttt agatgacttg 1380
tttaaagttc acaagcttaa gccaaatctg aagtggcgac aggcttttga cagctactta 1440
aaaactctgc ctccatacta cctattacca ttaaagaaag cactaaggat gatgggagct 1500
ccaaatctga tatcagataa tttagattgt ggacttagtt acagtgttat ctcttacctt 1560
aaaaaactca gccaacagac caaactagag tcagaacgaa tactagcatc agtggggaag 1620
aaacctcccc aggaaattgg aattaaagtg aaaaatcatt ctggaggtgg catgtccttg 1680
actcacaata aaaattttag aaaactattg aaagaaatca caggggaaac tgcacttaga 1740
ctgacagaat tgaacaccaa agaatttgct ggcttccaaa ttgggctctt aaacaaggta 1800
aattgtgaca taaatagttt gtatttgttt gaacttgtct gatacattta tgtctttcat 1860
attcctttag ttttgaaata atggacaatc ctattttgga gttagtagtt aataatggac 1920
cagtttagga attgtcagca gatatcaata tggaattaaa tggacaagat atgttttgtg 1980
ggtaacgccc atgaaagata aaggggagat gaagctggag taggcaggca gagcttccaa 2040
gctacaaagc tggcccatca cctgtgagag gaaggaggaa tgaaggagaa ttgggtgtga 2100
agaagctcag acactggtgt agcttggaga aagtcttggc cagcacagtg gggacctcta 2160
gcaccaagat ttcctacaga ggagggccac actgggcaga gctggccagg ccctaacacc 2220
ctgccattgt ctgttactgg ctatgggcag cctgggaagg gcatggcctc agctcagtgc 2280
tgcagcggat ccctatggca ccacagcagc gggaggctct ctgctaactg cactccttgc 2340
aggatcagcc attgctttcc tgaagggaga tcccagcagc tcacctcccc tccatggctg 2400
ccagaggaat acatctagag tacgagcagt ggtcctgcaa ggcggccaca ggtatcaatc 2460
tcgggatttg gtgttttta                                             2478

<210> 38
<211> 2928
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6715627CB1

<400> 38
ttcggctcga ggcgaggaca gggtacgtcg caggcttgtg cgggtcgggc tcggacctgc 60
gctgcctcgg gatgtaaagt ataacaagag ggtcgggatg gtcagcgtag gcctgtgtag 120
gcctgtcggg tgcccctgtc ccccatgctc cccccagcag ccggctccgc agtggtccac 180
tccggttgcc gggtgcggat tcgggttccg gaccgaaggc tgtgtgttct ccgccgttta 240
ttgtggcccc gacaggccgg ggttactgtg gcgaccacga gagcagcttt ggcgctatgg 300
aggagcccgg ggctacccct caaccgtatt tggggctgct cctggaggag ctacgcaggg 360
ttgtggcagc actgcctgaa ggtatgagac cagattctaa tctttatggt tttccatggg 420
```

```
aattggtgat atgtgcagct gttgttggat tttttgctgt tctctttttt ttgtggagaa 480
gttttagatc ggttaggagt cggctttatg tgggacgaga gaaaaagctt gctctaatgc 540
tttctggact aattgaagaa aaaagtaaac tacttgaaaa atttagcctt gttcaaaaag 600
agtatgaagg ctatgaagta gagtcatctt taaaggatgc cagctttgag aaggaggcaa 660
cagaagcaca aagtttggag gcaacctgtg aaaagctgaa caggtccaat tctgaacttg 720
aggatgaaat actctgtcta gaaaaagagt taaaagaaga gaaatccaaa cattctgaac 780
aagatgaatt gatggcggat atttcaaaaa ggatacagtc tctagaagat gagtcaaaat 840
ccctcaaatc acaagtagct gaagccaaaa tgaccttcaa gatatttcaa atgaatgaag 900
aacgactgaa gatagcaata aaagatgctt tgaatgaaaa ttctcaactt caggaaagcc 960
agaaacagcc tttgcaagaa gctgaagtat ggaaagaaca agtgagtgaa cttaataaac 1020
agaaagtaac atttgaagac tccaaagtac atgcagaaca agttctaaat gataaagaaa 1080
gtcacatcaa gactctgact gaacgcttgt taaagatgaa agattgggct gctatgcttg 1140
gagaagacat aacggatgat gataacttgg aattagaaat gaacagtgaa tcggaaaatg 1200
gtgcttactt agataatcct ccaaaaggag ctttgaagaa actgattcat gctgctaagt 1260
taaatgcttc tttaaaaacc ttagaaggag aaagaaacca aatttatatt cagttgtctg 1320
aagttgataa aacaaaggaa gagcttacag agcatattaa aaatcttcag actgaacaag 1380
catctttgca gtcagaaaac acacattttg aaaatgagaa tcagaagctt caacagaaac 1440
ttaaagtaat gactgaatta tatcaagaaa atgaaatgaa actccacagg aaaattaacag 1500
tagaggaaaa ttatcggtta gagaagaag agaaactttc taaagtagat gaaaagatca 1560
gccatgccac tgaagagctg gagacctata gaaagcgagc caaagatctt gaagaagaat 1620
tggagagaac tattcattct tatcaagggc agattatttc ccatgagaaa aaagcacatg 1680
ataattggtt ggcagctcgg aatgctgaaa gaaacctcaa tgatttaagg aaagaaaatg 1740
ctcacaacag acaaaaatta actgaaacag agcttaaatt tgaacttta gaaaaagatc 1800
cttatgcact cgatgttcca aatacagcat ttggcagaga gcattcccca tatggtccct 1860
caccattggg ttggccttca tctgaaacaa gagcttttct ctctcctcca actttgttgg 1920
agggtccact cagactctca cctttgcttc caggggggagg aggaagaggc tcacgaggcc 1980
cagggaatcc tctggaccat cagattacca atgaaagagg agaatcaagc tgtgataggt 2040
taaccgatcc tcataggggct ccctctgaca ctgggtctct gtcacctcca tgggaccagg 2100
accgtaggat gatgtttcct ccgccaggac aatcatatcc tgattcagcc cttcctccac 2160
aaaggcaaga cagatttgt tctaattctg gtagactgtc tggaccagca gaactcagaa 2220
gttttaatat gccttctttg gataaaatgg atgggtcaat gccttcagaa atggaatcca 2280
gtagaaatga taccaaagat gatcttggta atttaaatgt gcctgattca tctctccctg 2340
ctgaaaatga agccactggc cctggcttgt tcctccacct cttgctccca atcagaggtc 2400
cattgtttcc agtggatgca agaggcccat tcttgagaag aggacctcct ttcccccac 2460
ctcctccagg agccatgttt ggagcttctc gagattattt tccaccaggg gatttcccag 2520
gtccaccacc tgctccattt gcaatgagaa atgtctatcc accgaggggt tttcctcctt 2580
accttccccc aagacctgga tttttcccccc caccccaca ttctgaaggt agaagtgagt 2640
tcccctcagg tttgattcca ccttcaaatg agcctgctac tgaacatcca gaaccacagc 2700
aagaaacctg acaatatttt tgctctcttc aaaagtaatt ttgactgatc tcattttcag 2760
tttaagtaac tgctgttact taagtgatta cactttggct caaattgaag cttaatggga 2820
attataattc tcaggatagt atttgtaaat caagatgatt tacatatgaa tcttatgagt 2880
aaattattcc aattttattt tagacggtat actatttcac ttgattaa 2928
```

```
<210> 39
<211> 3486
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7727886CB1

<400> 39
gtcgccgggc gggccaggag cagcgcggac ccgagccggg cagggggcgc ccgccacggc 60
acccgcgcgc tcctagcgcc ccagacccgc ctgcgggccc ggatcctcct tgccactgtc 120
ccacccgccg tcctgccccc tccacccttt gtgtcgccgc agcccggtgc ccccggctct 180
gcgggacccc ggccgggccg gaccctggca aagcgccagg ccccgcgtgg gctcccggcg 240
agcggttgat ggcgagggg cgcggcgcgg gctctgtagc ccgagttccc gacgctggag 300
gcccggcccg cctcagccgc attgtcccgg gccgcgcgca ccggccctga gctgcgccgc 360
cgcagcaccc gcccgccgcc gcgggggcca tgcggagagc cgccgggatg gaggacttct 420
ccgcggagga agaggagtcc tggtacgacc agcaggacct ggagcaggac ttgcacctag 480
ctgcggagct ggggaagact ctgctggaga ggaacaagga gctggagggg tccctgcagc 540
agatgtactc caccaatgag gaacaggtgc aggagatcga gtacctaacc aagcagctgg 600
```

```
acacgctgcg gcacgtgaac gagcagcacg ccaaagtcta tgagcagctg gacctgacag 660
cccgggacct ggagctgacc aaccacaggc tggtgctgga gagtaaggct gcccagcaga 720
agatccatgg gctgacggag accattgagc gcctccaggc tcaggtggag gagctgcagg 780
cccaggtgga gcaactgaga ggcctggaac agctgcgagt gctccgggag aagcgggaac 840
gcaggcgtac catccacacc ttcccctgcc tcaaggagct gtgcaccagc ccccggtgca 900
aggatgcttt ccgcctacac agttcctccc tggagctggg cccgcggccc ctggagcagg 960
agaacgagcg gctgcagacc ctggtggggg cgctgcgctc ccaggtgagc caggagcggc 1020
agcgcaagga gcgggcggag cgcgagtaca ccgcggtgct gcaggagtac tcggagctgg 1080
agcgccagct gtgcgagatg gaggcctgtc gcctgcgtgt gcaggagctg gaggccgagc 1140
tgctggagct gcagcagatg aagcaggcca agacctacct actgggtccg gacgaccacc 1200
tggccgaggc cctgctcgca cccctcacgc aggcccctga ggccgacgat ccccagcccg 1260
gccgcgggga cgacttgggc gcccaggacg gggtctcctc accggtcagcc tctccaggcc 1320
acgtggtgcg caagagctgc agcgacactg cgctcaacgc catcgtggcc aaagacccag 1380
ccagccggca cgcgggcaac ctcacactgc acgccaacag cgtgcgcaag cggggcatgt 1440
ccatcctgcg ggaggtggac gagcagtacc acgcgctgct ggagaagtac gaggagctgc 1500
tgagcaagtg ccggcagcac ggggccggag tgcggcacgc cggcgtgcag acctcgcgcc 1560
ccatctcccg ggacagctcg tggagggacc tgcgcggggg tgaggagggc cagggtgagg 1620
tcaaggcagg agagaagagc ctgagccagc acgtggaggc cgtggacaag cggctggaac 1680
agagccagcc cgagtacaag gcgctcttca aagagatctt ctccaggatc cagaagacca 1740
aggctgacat caacgccacc aaagtcaaga cgcacagcag caagtgaccc ttctccggcc 1800
tgcagcctcc cccagggtgg aagccgtggg gtccctcagg cctgggcggt gcagcttcca 1860
gagagcgagc gccctttagc ggcctgccac cacagcacgc ggcctcctga tccggaagca 1920
cgcagcatgt tccctgctga gcggaggcag cccacctgtc ctgcctccca ggagcccttg 1980
gccacctcgc gccagcccaa aggcgcagct ctgagttcaa agccaaatgt ccccactacc 2040
ccagggatcc cccagctccc ccagcccctg gcttcctgac cctgcgcctc accctcagac 2100
tggtgaccag gcttctgaaa gccattctgg atcagttggc ttttttttttt ttttggtta 2160
agtttgtttt ttctaagaca tttgcaatgc aaggtctcct tgacccctttg ccacaactgg 2220
aaacacttga aaggggaccc cagggccagc tgtttcaggg gtttcctgga ccacccactg 2280
cttctcccca accctgatgc gctgacattc ccttagcacc agctgtccca cctccaggt 2340
cctgaccagg tcagagatgt cccctgccat gcagagcagg aagcctcagc tgggcctgga 2400
gtgtccctgc tccagccctg ccagggacag gtttctccct ggatactctt ggcccaccgc 2460
agatctgtag ccagtcagag gaggaggaga aggagccct cagcagagtg gtgcagtttc 2520
gctcagagct tgtctccttg gcttccaacc ccagaaatgc ctgctgggcc ttaagctttc 2580
caggggccgg ggcagtgggg agcccccatc ccttcacacc gccaccaact aaccaagctt 2640
ggcctctgac tcccgtctct gtgcttgccc ccatctcagg gaccatgatg tctcagtcac 2700
tccacgctcc ccacaggcca accctggcac aggtcatgtc tgcagcccc agaatgttct 2760
ggacatgcac caccagccgg tggtcccaat gtccacccct gcctccccctt cactggggac 2820
tggggttttc gccccatgct gcatcgtgtt gtattgggat ggggctgagg aacatgctcc 2880
ctccctccct cccataaaat gcctgctctt cacctcccac ctttgtgggg ggcttttgag 2940
gacccagctg cgtcaggagt tttgcttcaa gatgtcagaa agtcaagttc agctaagaga 3000
cacccaggtc cccagcttgc cctgagcagc ccttcagggc ttctggttcc ttctgccgcc 3060
cttcctgaga ccttagaaac cagaagagcc atacagtcag tggaggcggg gggggcccctg 3120
gcctctgcac cgggatccca gtgggaacct tcatgcctta tttatttcta atgggtaaag 3180
gggtttttctt accaagcatc cctgacctcc tggagacacc acctgctttc cgggcggcac 3240
tgtgatggga gctggtggcg actgagtcct tctgtacgtg caactgggaa acttttgtcc 3300
tttgaggcta ggcagctccc tgccctccgt gtgtgtctgt tatctggggg agaggagtgt 3360
ggaagggttg ggggaagagc tccagcctgt ctgctcccca gctctgtagt ggcagaccag 3420
cgtcaccttt gaagtatacg tgagagaaat atatttacaa atgctttatt ctcttcttta 3480
ataaaa                                                          3486
```

```
<210> 40
<211> 2535
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 914113CB1

<400> 40
agcgccggcg cggggccgcg agaacgcagc gcaggggagc agcccgaggc ggacaccgcg 60
agccgcccgg cactcccgca gtccagccgg ctcctctagc ccggccacgg ctccgctgcg 120
ggccacccag gattactcgc gtctggctcc aggcgccgag aaggcgcgct gggcgcccgt 180
```

```
ggccgccgcg ccagctcctc ctcctcccgc tgctcctgct cccggggcga gcgcgcagcc 240
ccgagcccgc cccgcgcctc ccggagccct cccccccgct gctcccatgc gcgcgggtgg 300
gtcatgagca cagcgccctc gctttctgcc ctaagaagca gtaagcacag cggcggcggc 360
ggcggcggag gcggaggcgg aggcggaggc ggcggtgcag accctgcctg gaccagcgcg 420
ctctctggaa atagctccgg ccccggccca ggctcgtccc cggccggcag caccaagcct 480
tttgtgcacg ccgtgccccc ctctgacccc ctgcgccagc caaccgcct gccaatcaag 540
gtgctgaaga tgctgacggc acgaactggc cacattttgc accccgagta cctgcagccc 600
ctgccttcca cgccggtcag ccccatcgag ctcgatgcca agaagagccc gctggcgctg 660
ttggcgcaaa catgttcgca gatcgggaag cccgaccct cgccctcctc caaactctcc 720
tcggttgcct ccaacggggg cggcgcgggc ggtgccggag gcggtgctgc gggcgacaag 780
gacaccaaat cgggcccct gaagctgagc gacatcggcg tggaggacaa gtcgagtttc 840
aagccgtact ccaaacccgg ctcggataag aaggagccgg gaggcggcgg tggaggcggt 900
ggcggtggcg ggggcggcgg cgggggtgtt tcgtcggaga agtcgggatt ccgggtaccg 960
agcgccacct gccagccatt cacgcccagg acaggcagcc ccgagctcca gcgcctcggc 1020
ctgctcgccg ggaggtatgc tgtcctcggc cggggtgcc ccggagggca aggacgacaa 1080
gaaagacacc gacgtgggcg gccggtggca agggcaccgg ggggcgcctc ggccgaaggg 1140
ggacccacgg ggctggcaca cggccggatt agctgcggcg gcgggattaa tgtggatgtg 1200
aaccagcatc cggatggggg cccgggaggc aaggctctgg gctcggactg cggcggttca 1260
tcgggctcca gctccggctc cggccccagc gcgcccacct cctcctcagt gttgggctct 1320
gggctggtgg ctcccgtgtc accctacaag ccgggccaga cagtgttccc tctgcctccc 1380
gcgggtatga cctacccagg cagcctggcc ggggcctacg ccggctaccc gccccagttc 1440
ctgccacacg gcgtggcact tgaccccacc aagccgggca gcctggtggg ggcgcagctg 1500
gcggcggccg cggccgggtc tctgggctgc agtaagccgg ccggctccag ccctttggcc 1560
ggagcgtctc cgccgtccgt gatgacagcc agtttgtgcc gggaccctta ctgcctcagc 1620
taccactgcg ctagccacct ggcaggggcg gcggccgcca gcgcttcttg cgcacatgat 1680
ccggctgctg cggctgcggc gctgaagtcc ggatacccgc tggtgtaccc cacgcacccg 1740
ctgcacggtg tgcactcctc gctaacggcc gccgcggctg ctggcgccac accgccctcc 1800
ctggccggcc accccctcta cccctacggg tttatgctcc ctaacgaccc actcccccac 1860
atctgcaact gggtgtcggc caacgggccg tgcgacaagc gcttcgccac gtccgaagag 1920
ctgctgagcc acttgcggac ccatacggca tttcccggga cagacaaact gctgtcgggc 1980
tacccagct cgtcgtctct ggccagcgct gccgcggccg ccatggcttg ccacatgcac 2040
atccccacct cgggcgcacc gggcagccct gggacgctgg cgctgcgcag cccccaccac 2100
gcgctgggac tcagcagccg ctaccacccc tactccaaga gcccgcttcc cacgcctggc 2160
gcccccgtgc cggtgcccgc cgccaccgga ccgtactact cccccctacgc cctctacgga 2220
cagagactga ccaccgcctc ggcgctgggg tatcagtgag ggcggccggg agggcgagcg 2280
agggagagga gggagagggg gaggggagga gtccaggag aggcgggatc acggcccagg 2340
ctgctgacac ccgcgcgtgg ggaggactcg gccacgaaa ggaaagaaat gtataccgta 2400
tctatctacc cgacagcagc gaccgagacc cggtgggaca ctcccccttct ccccactttc 2460
acctccccac ccaaacttta taaagttga aaaatatca tttgactttt tatagaaaaa 2520
aaaaaaaaaa aaaaa                                                2535
```

```
<210> 41
<211> 5514
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1953711CB1

<220>
<221> unsure
<222> 208
<223> a, t, c, g, or other

<400> 41
cggccgccgg ccgcagcggg ctgagattgt tgtcctctgt caccagggcg gctgggctcc 60
cgcagtcctg cagaccgcgc ccgatcccgg cgacagggcg ggcggacagc cgcgcatccc 120
cggggtcccg ccgagcctgg gcgcagagag ccgggaggaa gcgttcgctc gcttcgcctt 180
gctgctggga aactgaacga agggcccnga gagagaaggt tcttgagttc atgtaagagg 240
acagtcttaa aacggaagaa gaaagagaag cagttcagtc tttgggagag ctgcctcctt 300
gttgagtgct gcaaaggcct ggaattcatt tatgacagaa tagatctaga aaagtccaag 360
catgtttttct agagtggtgt agccctgtgc tgcctccagt gaagagtctc ttggtgttgg 420
```

```
cttcgtgctt ccggagggac catggcaacc tccagagggg cctcccggtg tcctcgggat 480
atcgccaatg tgatgcagag gctgcaagat gagcaagaga tagtacaaaa acgaactttc 540
acaaaatgga tcaactctca tctggccaag cggaaacctc caatggtggt ggacgatctt 600
tttgaagaca tgaaagatgg tgttaaactg cttgcccttc tggaggtcct gtctgggcag 660
aaactgcctt gtgaacaagg acgccggatg aagcgaatcc atgctgtggc taacattggc 720
acggcactca agttcctcga aggaagaaag attaaattag tcaacattaa ctccaccgat 780
atagctgatg gccgaccctc aatagttctt ggattgatgt ggaccattat tctatatttc 840
cagattgaag agttgaccag caacctgccc cagctccagt ctttgtccag cagcgcatcc 900
tccgtggaca gcatagttag ctctgagact cccagcccac caagtaaacg gaaggtgacc 960
accaagatcc aaggaaatgc taagaaggct ttattaaagt gggttcagta cacagctggc 1020
aagcagactg gaatagaagt aaaagatttt gggaagagtt ggagaagcgg ggttgccttt 1080
cattcagtta ttcatgccat tcgaccggaa ttggtggact tggagacagt gaaaggcaga 1140
tccaaccgag aaaatttgga ggatgctttc actatcgccg aaacagaact ggggatccca 1200
agactgctag atcctgaaga cgttgatgtg gataaaccag atgagaaatc tattatgacc 1260
tatgtagccc agtttctgaa acattatcct gacatccaca atgcaagcac tgatgggcaa 1320
gaggatgatg aaatacttcc aggtttccca tcttttgcaa attctgtaca aaattttaag 1380
agagaagaca gagtaatttt taaggaaatg aaagtttgga tagaacaatt tgagagagat 1440
ttgacaagag cacagatggt ggaatcaaat ttacaggata aatatcagtc atttaagcac 1500
ttcagagttc aatatgaaat gaagaggaaa cagattgaac atttaataca accattacac 1560
agagacggta aattgtcact tgaccaagca ttggtaaaac aatcttggga tagagtgacc 1620
tccaggctct ttgactggca tatacagctt gataaatctc ttcctgcacc tctgggcacc 1680
ataggtgcct ggctgtacag agcggaggtg gccctgagag aggaaataac cgttcaacag 1740
gtccacgagg aaacagcaaa cacgatacaa cggaaacttg agcaacataa ggatctgctt 1800
caaaacacgg atgcccacaa aagagcattc catgaaatct accggaccag gtctgttaac 1860
gggattccag tgccacctga tcaattagag gacatggccg agaggtttca ttttgtttcc 1920
tccacatcag agctacacct aatgaaaatg gaatttttag aattaaagta ccgtctgctc 1980
tcactgctgg ttcttgcaga gtcaaagctg aagtcttgga tcattaagta cgggaggaga 2040
gagtcagtgg agcagcttct acaaaactac gtgtctttta tagaaaatag caagttcttt 2100
gaacaatatg aggtgacata ccagatcttg aaacagacag ctgagatgta tgtcaaagca 2160
gatggttcag tggaagaagc tgagaatgtg ataaaattca tgaatgaaac caccgctcag 2220
tggaggaatc tctcagtaga agtgaggagt gtgaggagca tgctggaaga agtgatctct 2280
aactggggatc gctatggcaa tacagtggct agtctgcaag cctggctaga ggatgctgaa 2340
aaaatgctca tcaatcaga aaacgccaaa aaggattttt ttcgaaattt acctcattgg 2400
attcagcagc atactgccat gaacgatgct ggcaatttc taattgaaac ctgtgatgag 2460
atggtttccc gtgacctgaa gcagcaatta ctgttgctaa atgggcggtg gaggagttg 2520
tttatgggag tcaagcaata tgctcaagct gatgagatgg acagaatgaa gaaggaatac 2580
acagactgtg ttgttaccct gtctgctttt gcaacggaag cccataagaa actttctgaa 2640
cccttagaag tctcttttat gaatgtcaag ctattaattc aagacttgga ggatattgag 2700
cagagggtgc ctgtgatgga tgcccaatac aagataatta caaagacagc acacctcatt 2760
accaaagaaa gcccccaaga agaaggaaaa gaaatgtttg cgaccatgtc aaagctcaaa 2820
gagcagctaa ccaaggtcag agaatgttac tccccactcc tttatgagtc tcagcagctg 2880
ttgattccgt tggaggaatt agaaaagcag atgacgtcct tttatgactc acttgggaaa 2940
atcaatgaaa ttatcacagt tcttgagcgt gaggcacaat cgagtgccct ttttaaacaa 3000
aaacatcagg aactgttagc ttgtcaagaa aactgtaaga aaaccttgac acttattgag 3060
aaaggcagtc aaagtgttca aaagtttgtg accttgagca acgtgttaaa gcattttgat 3120
cagacgaggc tacaaagaca gattgcagat attcatgttg cttttcagag tatggtaaag 3180
aaaactggag attggaagaa gcatgtggaa accaacagtc gcttgatgaa gaagtttgag 3240
gagtctcgag cagagttgga gaaggtactg cggattgctc aggagggcct ggaggaaaag 3300
ggggatccag aggagctcct gcggagacac actgagtttt tcagtcagct ggatcagagg 3360
gtgctcaatg ctttcctgaa agcttgtgat gaactcaccg acatccttcc agagcaggag 3420
cagcaggggc tgcaggaagc tgttcgaaag ctcccacaaac aatggaagga tcttcaagga 3480
gaagcccctt atcatttgct tcatctgaag attgatgtgg agaagaatag gttcttagcc 3540
tctgtagaag aatgcagaac tgagctggat cgagagacca agctgatgcc ccaggaaggc 3600
agtgaaaaga taattaaaga gcacagggtt ttcttcagtg acaaaggtcc tcatcatctc 3660
tgtgagaaaa ggttacagct catcgaggaa ctctgtgtga aactcccagt gcgggaccca 3720
gtaagggaca cacctggaac ctgtcacgtg actctcaaag agctcagagc tgccattgac 3780
agcacctaca ggaagctcat ggaagaccca gacaagtgga aggactacac tagcagattc 3840
tctgagttct catcttggat atctacaaat gagacacaat aaaggggat caagggtgag 3900
gccatcgata ctgccaacca cggagaggtt aaacgtgccg ttgaagagat cagaaatggt 3960
gttaccaaaa ggggtgagac cctcagctgg ctgaaatcca ggctgaaagt tttgacagaa 4020
gtttcttctg agaatgaagc ccaaaagcag ggagatgagc tggcaaaatt atccagctct 4080
ttcaaggctc ttgtgacgct gctgtcagag gttgaaaaga tgctaagcaa ttttgggggac 4140
tgtgtccagt acaaagaaat agtcaaaaat tctctcgaag aattaatttc tggctctaaa 4200
```

```
gaagtccagg aacaagctga gaagatcttg gatactgaaa atctgtttga agcacagcag 4260
ttacttcttc atcaccagca aaagacaaag cggatctcag caaagaagag agatgtgcag 4320
cagcagatcg cgcaggcgca gcagggagaa ggggggctgc ctgaccgagg ccacgaggag 4380
ctgcggaagc tggagagcac actggatggc ctggagcgca gccgggagag gcaggaacgc 4440
cgcatccagg tcacattaag aaaatgggag cgatttgaaa caaacaaaga aacagtagta 4500
agataccttt ttcaaacagg ttccagtcat gaacgcttct tgagttttag cagtttggaa 4560
agtttatctt cagaactgga acaaacaaag gagtttttcta aacggacaga aagtattgca 4620
gtccaggctg agaaccttgt aaaggaagct tcagagatac cgcttgggcc ccaaaataag 4680
cagctgcttc aacagcagac caagtcaatc aaagaacaag tcaaaaaatt agaagacacg 4740
cttgaagaag agtatgtgat tgacaagtcc taaacttttct tctctgagat aaagtttcat 4800
acaatctttc ctgtaccttg tattcaaaac actcttttaa aatctcaaag tgtctgtgta 4860
tttcagcatg ttttgaggaa acaactcaca gttcaaaaga aagtatcgct aatacagaaa 4920
ccaatatcta taacagagcc caaaaaatat aaaggatgtg ggttttgcat cttaaactga 4980
tcatgttcat gagaaagcca tatctattct attctgtggc ctttgtacat tgtagaggga 5040
atcttgaaaa agaactaata tttaaaataa ttttttttact atattattct gctgtcacca 5100
tttagagcga aaaggagata ttttgttagt gtagattcca ggcctaaata cacatcacat 5160
agaccatata tctccaacct gaagaagctc ctggagcttg tttacagtgc ctcggtattc 5220
aagttatcct gactaatatg ctctttccag aaattaactt taaaatattt tatttttaaa 5280
ctttttaatgt ttgtttatct gaaaaaaaaa aaactgcttg gcctattgat atcctgttat 5340
atacccaaaa tatacaaaaa agttgggagg ggaaaaacat atgaggatac agccaatatt 5400
attcaaaaca atatggctaa acatttctat ttttaacata atttagtggg caaatcaatg 5460
atacataata gcatattatt attaagggac aggccatcat ttctaaggtg gcgt       5514
```

<210> 42
<211> 3715
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1595275CB1

<400> 42
```
ggtcgctggc agctgagcgg cggaaccctt cgctggcggg agcgcggttc ggggccgcag 60
aggggacatt gaataaacct gagtattgtg tgcctgtatt ttgactgcga cctatcagat 120
gagatcaggt caagagacaa ctgcctttga aagagagtga gtgactgttt ccaaaacagc 180
aagcataggg agagccgccg gaagtattga cggaaagctt gaattaccta catgattgat 240
gatggtctcc agctgtcaaa gaagattgtg gttgccaaga gaaaggtaac catttctcat 300
ccccttgagt ggaattggat gcctctagga gcccacattg gctggcagta gacatggccg 360
aatttacaag ctacaaggag acggcctcca gccgccactt gcggtttaag ctacagagtc 420
taagccgccg ccttgatgag ttggaggaag ccacaaaaaa cctccagaaa gcagaggatg 480
aactcctgga tctccaggac aaggtgattc aggcagaagg tagcaactcc agcatgttgg 540
cggagattga agtgctgcgc cagcgggtgc tgagaattga aggcaaagac gaggaaatta 600
agagagcaga ggatctgtgt cgtctgatga aggagaaact tgaagaggaa gaaaacctca 660
cccgggagct gaaatctgag attgagcggc ttcagaaacg aatggctgaa ttagagaagc 720
tagaagaggc cttcagcagg agtaagaatg actgtaccca gctgtgtctg agcctgaatg 780
aggagagaaa tctgaccaag aaaatctcct ctgagctgga aatgctcaga gtcaaagtga 840
aagaactaga atcttctgag gaccgcctgg ataaaactga gcagagttta gcgtcagagt 900
tagaaaagct gaaatcatta actctgagct ttgtaagtga gagaaaatac ttgaatgaaa 960
aggagaaaga aaatgagaaa ttgataaaag aactcactca aaaactagag cagaacaaaa 1020
aaatgaaccg agattataca aggaatgctt ctaatctgga aagaaatgac ctacggattg 1080
aggatggcat ctcttccaca ctgccgtcca aagaatcaag aaggaagggt ggtctggact 1140
acctaaagca ggtagagaat gaaacaagaa acaaatcaga aaatgaaaag aaccgcaatc 1200
aggaagacaa caaagtcaaa gaccttaacc aagagattga gaaacttaag acgcaaatca 1260
aacactttga atcgttggaa gaagagctta agaaaatgaa gtccaaaaat aacgaccttc 1320
aggataatta cctaagtgaa caaaataaaa acaaattatt agccagccaa ctggaggaga 1380
taaagctaca aatcaagaaa cagaaagagt tagaaaatgg agaggtagaa ggggaagatg 1440
ctttcctgtc cagcaaaggc agacatgaga ggactaagtt tagaggccac ggaagtgaag 1500
cttctgtgtc caagcacaca gcgcgggaac tgtctcctca gcataagcgg gaacgactcc 1560
ggaacaggga gtttgctctc aacaatgaaa actattctct gagcaacagg caggtttctt 1620
ctcccagttt caccaacagg agggcagcaa aagcttctca catggggggtg agtacagaca 1680
gtgggactca ggagacaaag aaaactgaag accggtttgt acccagctcc tccaagagcg 1740
aagggaagaa gtctagggag cagccctcag tgctgagtcg ctaccccccg gctgctcagg 1800
```

```
agcacagtaa agcgtggaag gggacttcca agccaggcac cgagagtgga ctgaagggaa 1860
aagtggagaa gacaacacga acgtttagtg acaccaccca tggatctgtt cccagtgacc 1920
cattgggtag agctgacaag gcttctgaca cctcctctga gactgtcttt ggcaagaggg 1980
gacacgtgct tggcaacgga agtcaagtaa ctcaggctgc aaactctggc tgttctaagg 2040
ccattggagc cctggcctca tctcgaagat cctcctcaga agggctctct aaaggcaaaa 2100
aggctgccaa tggccttgag gctgataaca gttgcccgaa ttccaaggct cctgttctat 2160
cgaagtatcc ttatagctgt agaagccaag agaacatcct tcagggattt tcaacctcac 2220
ataaagaagg ggttaatcaa cctgcagcag ttgtgatgga agacagcagt ccgcatgaag 2280
ccttgaggtg tcgagtcatc aaatccagtg gcagagagaa gccagactca gatgatgact 2340
tggacatagc atctcttgtt actgccaagt tggtaaatac aaccatcact ccagagccag 2400
agcccaaacc acagcctaac tctagagaaa aggctaaaac ccgaggggca cctagaacct 2460
ccctatttga gaatgataaa gatgctggaa tggagaatga atctgtgaaa tctgtcagag 2520
cctccaccaa taccatggag ctcccagata ccaatggtgc tggggtaaaa agccaaaggc 2580
cctttagccc cagagaggcg ttgcggtcta gagccatcat caaacctgtt attgttgata 2640
aggatgtgaa aaaaatcatg ggaggatctg gaacagagac tacgttggag aaacagaaac 2700
ccgtctccaa accagggcca aacaaagtga caagtagcat tactatatat ccatctgaca 2760
gcagcagccc cagagctgct ccgggtgagg ccctgaggga gaggcacaca tccactagca 2820
atatccaggt ggggctggca gagctcacat cagttagcaa ccatgtcagc tccccttttg 2880
agctctccat tcacaaacat gacatcaccc tgcagcttgc agaagcggag agaatggcag 2940
atgggcccct gaagaacagg ccagaaacag tggtctctcg gagcagcatt ataatcaagc 3000
catcggatcc tgtggagagg aatagccatg cacctccagc ggagacaatc aggtggaaaa 3060
gccatagtgc cccttcagaa gtgggcttct cagatgccag acatgttact gtgcggaatg 3120
cctggaagag taggcgagac ttgaaatctt tagaagaccc cccaactcga ataggtaaaa 3180
acgtggaatc taccaatagc aatgcctaca cccagaggtc ttccacagac ttctcagaac 3240
tcgaacagcc caggtcctgc cttttgagc agggcactcg aagggtagga ccaagttcag 3300
gggatgcccc tgagccctcc tccagaagga cccaaagtag cctcactgtg tcagaggtgc 3360
ttacccgtcg gaatcgggta ggagacacca tcactgtcgc agcctggaac cactcagcaa 3420
gcatggagga agaaggggaa gactgtacac tcagtgtcta caggcaactg cacaatccct 3480
ggatccgtct gaactgcctg ggaagcaggg gctgccagag tctgggcgag tacgggcgag 3540
ggaacgatta cgggccaaac aggccctgtg ctgaggagac actgagcccg catgggtgtc 3600
tgttgtctgc cctctgactc ctgcttctca ggatcttcca ccgtactgcc ccacgaagga 3660
tcaaggccag ttaaccaggc aagacacaag cctttgctgg gaaaactgtg gtaga      3715
```

```
<210> 43
<211> 1082
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 72332548CB1

<400> 43
ccgcggacga agatggcgac cgccatgtac ttggagcact atctggacag tatcgagaac 60
cttccctgcg aacttcagag gaacttccag ctgatgcgag agctggacca gaggacggaa 120
gataagaaag cagagattga catcctggct gcagagtaca tctccacggt gaagacgctg 180
tctccagacc agcgcgtgga gcgcctgcag aagatccaga acgcctacag caagtgcaag 240
gaatacagtg acgacaaagt gcagctggcc atgcagacct acgagatggt ggataaacac 300
attcgaaggc ttgatgcaga cctggcgcgc tttgaagcag atctgaagga caagatggag 360
ggcagtgatt ttgaaagctc cggagggcga gggttaaaaa aaggccgggg tcagaaagaa 420
aaaagagggt cccggggccg aggcaggagg acatcagagg aagacacacc aaagaaaaag 480
aagcacaaag gagggtctga gttcactgac accatcctgc ccgtgcaccc ctctgatgtg 540
ctggacatgc ccgtggaccc aaacgaaccc acgtactgcc tgtgccacca ggtctcctat 600
ggggagatga ttggctgtga caatccagac tgtccaattg agtggtttca ctttgcctgc 660
gtggacctta ccacgaaacc caaaggaaaa tggttctgtc acggtgtgt ccaggaaaag 720
aggaagaaga agtaggagga gcatgtgtgc ccaggatccg aggagcaagt taatctgtcc 780
cttcattcgt gtcgcaatat ttcccttcct tttaaaacta ccttgttcgg ttgatactta 840
gtaactccgt ggccagttga agcgctggat gcttcctaga acaagaacca ccaaagcctg 900
ttcgcacaga agggcgacct tgcagggact cgccgccgcg acctcagtgt ggcttttaca 960
ggacatccac cccgagcatc agcagggacc ccggcggacg tgggcaggc gcgcgtgagc 1020
tcgggctgcc cggccggggcg tgcggcgggg gacatggtaa cctggtccac ggagggcggc 1080
ca                                                            1082
```

```
<210> 44
<211> 2621
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7322834CB1

<400> 44
gcgccgagcg ccgcgccgcc ccgggcccct ccgccgccgc cggcccggac atggccgcca 60
acatgtacag ggtcggagac tacgtctact ttgagaactc ctccagcaac ccatacctga 120
tccggagaat cgaggagctc aacaagacgg ccaatgggaa cgtggaggcc aaagtggtgt 180
gcttctaccg gaggcgggac atctccagca ccctcatcgc cctggccgac aagcacgcaa 240
gggaaataga agaggaaatg gagaatccgg aaatggtgga cctgcccgag aaactaaagc 300
accagctgcg gcatcgggag ctgttcctct cccggcagct ggagtctctg cccgccacgc 360
acatcagggg caagtgcagc gtcaccctgc tcaacgagac cgagtcgctc aagtcctacc 420
tggagcggga ggatttcttc ttctattctc tagtctacga cccacagcag aagaccctgc 480
tggcagataa aggagagatt cgagtaggaa accggtacca ggcagacatc accgacttgt 540
taaaagaagg cgaggaggat ggccgagacc agtccaggtt ggagacccag gtgtgggagg 600
cgcacaaccc actcacagac aagcagatcg accagttcct ggtggtggcc cgctctgtgg 660
gcaccttcgc acgggccctg gactgcagca gctccgtccg acagcccagc ctgcacatga 720
gcgccgcagc tgcctcccga gacatcaccc tgttccacgc catggatact ctccacaaga 780
acatctacga catctccaag gccatctcgg cgctggtgcc gcagggcggg cccgtgctct 840
gcagggacga gatggaggag tggtctgcat cagaggccaa ccttttcgag gaagccctgg 900
aaaaatatgg gaaggatttc acggacattc agcaagattt tctcccgtgg aagtcgctga 960
ccagcatcat tgagtactac tacatgtgga agaccaccga cagatacgtg cagcagaaac 1020
gcttgaaagc agctgaagct gagagcaagt taaagcaagt ttatattccc aactataaca 1080
agccaaatcc gaaccaaatc agcgtcaaca acgtcaaggc cggtgtggtg aacggcacgg 1140
gggcgccggg ccagagccct ggggctggcc gggcctgcga gagctgttac accacacagt 1200
cttaccagtg gtattcttgg ggtccccta acatgcagtg tcgtctctgc gcatcttgtt 1260
ggacatattg gaagaaatat ggtggcttga aaatgccaac ccggttagat ggagagaggc 1320
caggaccaaa ccgcagtaac atgagtcccc acggcctccc agcccgggagc agcgggagcc 1380
ccaagtttgc catgaagacc aggcaggctt tctatctgca cacgacgaag ctgacgcgga 1440
tcgcccggcg cctgtgccgt gagatcctgc gcccgtggca cgctgcgcgg cacccctacc 1500
tgcccatcaa cagcgcggcc atcaaggccg agtgcacggc gcggctgccc gaagcctccc 1560
agagcccgct ggtgctgaag caggcggtac gcaagccgct ggaagccgtg cttcggtatc 1620
ttgagaccca cccccgcccc cccaagcctg accccgtgaa aagcgtgtcc agcgtgctca 1680
gcagcctgac gcccgccaag gtggcccccg tcatcaacaa cggctccccc accatcctgg 1740
gcaagcgcag ctacgagcag cacaacgggg tggacggtct ggcaaaccac ggacagacca 1800
ggcacatggg accaagccgg aacctcctgc tcaacgggaa gtcctacccc accaaagtgc 1860
gcctgatccg ggggggctcc ctgcccccag tcaagcggcg gcggatgaac tggatcgacg 1920
ccccggatga cgtgttctac atggccacag aggagaccag gaagatccgc aagctgctct 1980
catcctcgga aaccaagcgt gctgcccgcc ggccctacaa gcccatcgcc ctgcgccaga 2040
gccaggccct gccgccgcgg ccaccgccac ctgcgccgt caacgacgag cccatcgtca 2100
tcgaggacta ggggccgccc ccacctgcgg ccgcccccg ccctcgccc gcccacacgg 2160
ccccttccca gccagcccgc cgcccgcccc tcagtttggt agtgccccac ctcccgcccct 2220
cacctgcaga gaaacgcgct ccttggcgga cactggggga ggagaggaag aagcgcggct 2280
aacttattcc gagaatgccg aggagttgtc gtttttagct ttgtgtttac tttttggctg 2340
gagcggagat gaggggccac cccgtgcccc tgtgctgcgg ggcctttttgc ccggaggccg 2400
ggccctaagg ttttgttgtg ttctgttgaa ggtgccattt taaattttat ttttattact 2460
tttttttgtag atgaacttga gctctgtaac ttacacctgg aatgttagga tcgtgcggcc 2520
gcggccggcc gagctgcctg gcgggggttgg cccttgtctt ttcaagtaat tttcatatta 2580
aacaaaacaa gaaaaaaaat cttataaaaa ggaaaaaaac c              2621

<210> 45
<211> 5539
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1925714CB1
```

<400> 45

```
aaaccaaaac ctgaatccag cagaagaaag aaaagaacca agattagagc agaactaaat 60
gaaattgaaa caaaaaaaat agaaaaaata aataaaacaa aaagctgttt tttgaaataa 120
tacataaaat tgatagacca ttagaaaaat taaccgagaa aagaagagag aggattaaaa 180
taagctcaat tagaaacaaa atgggagata ttacaaccaa caccacagaa atacaaaaga 240
tcactaaagg ctactatgaa cacctttatg tgcataaact agaataccta gaggagatga 300
ataaattcct gaaaaaatac aatccttcta gcttaaatca gaaagaatta gaaatcctga 360
acagaccaat aacaagcagc cagattgaac tggtaattaa gaaattacca acaggctgga 420
ctccacctct gggggcaggg catagccaaa caaaaggcag cagaaacctc tgcagactta 480
aatgtccctg tctgacagcg gtgaagagaa tagttgttct cccaggacgc agctggagat 540
ctgagaacgg acagactgcc tcctcaaatg ggtccctgac cccaattcca gaggaacgat 600
caggcagcaa catttgctgt tcaccaatat cagctgttct gcagcctctg ctgctgatac 660
ccaggcaaac agggtctgga gtggacctcc ggcaaactcc aacacgtctg cagctgaggg 720
ttctgactgt tagaaggaaa actaacaaac agaaaggaca tccacaacca taccccatct 780
gtacatcacc atcatcaaag aacaaagaga agtccttaaa ggacctgatg gagctgaaaa 840
ccacggcacg agaactacat gatgaatgca caagcctcag tagacgattt gatcaactgg 900
aagaaagggt atcagtgatg gaagatcaaa tgagtgaaat gaaaataaaa tacctaggaa 960
tccaacttac aagggatgtg aaggaactct tcaaggagaa ctacaaacca ctgctcaatg 1020
aaataaaaga agatacaaac aaatggaaga atattccatg ctcatgtgta gaaagagtca 1080
atatcgtgaa aatggccata ctgcccaagg taatttatag attcaatgcc atcaccatca 1140
agctacaaat tactttcttc acagaattgg aaaaaactac tttaaagctc atatggaacc 1200
aaaaaagagc tcacattgcc aagtcaatcc taagccaaaa gaacaaagct ggaggcatca 1260
cgctacctga cttcaagcta tactacaaga ctacagtaac caaaacagca tggtactggt 1320
accaaaacag aggtacagac caatggaaca gaacagagcc ctcagaaata atgctgcata 1380
tctacaacta tctgatcttt gacaaacctg acaaaaacaa gaaatgtgga aaggattccc 1440
tatttaataa atggtgctgg gaaaactggc tagccatatg gagaaagctg aaactggatc 1500
ccttccttac accttacaca aaaattaatt ctagatggat taaagactta aatattagac 1560
ctacaaccat aaaaacctta gaagaaaacc taggcaatac cattcaggac ataggcatgg 1620
gtaagaactt catgtctaaa acaccaaaag cagtagcaac aaaagccaaa atttacaaat 1680
gggatctaaa taaactaaag agcttctgca ctgcaaaaga aattaccatc agagtgaaca 1740
ggcaacctac aaaatgggag aaaattttg caatctactc atctgacaaa gggctaacat 1800
ccagaatctg caatgaactc aaacacattt acaagaaaaa aacaaacaac cccatcaaaa 1860
agtgggtgaa ggatatgaac agacacttct caaaagaaga catttatgca gtcaaaagac 1920
atgaaaaaat gttcatcatc actggccatc agagaaatgc aaatcaaaac cacaatgaga 1980
agacatggac acaggaaggg aacatgacac accgggggcc tgttgtgggg tggggggagg 2040
gtggagggat agcatttgga gatcgcgggc ctgtcggtgt ctgctcctac acgccaacgc 2100
cggtgggcag gaccatgtct ctggtaagcc agaattcgcg ccgccgccgc cgccgcgttg 2160
caaaggctac tgcgcacaac agcagctggg gcgaaatgca ggccctaat gccccggtc 2220
tccccgctga tgtgccaggc tcagacgtcc cccagggtcc cagcgattcc cagatcctcc 2280
agggcctctg cgcctctgag ggcccaagca cctccgttct gcccacctcc gctgagggcc 2340
caagcacctt tgtgccgccc accatctctg aggcctcaag cgcctccggg cagcccacca 2400
tctctgaggg acctggcacc tccgtgctgc ccaccccag tgaggccta agcacctccg 2460
ggcctcccac catctctaag gggctgtgca cctctgtgac gcttgccgcc tctgagggcc 2520
ggaacacctc caggccgccc acttcctctg aggaacctag cacctccgtg ccgcccaccg 2580
cctctgaggt accgagcacc tccctgccgc ccacccctgg tgagggaacg agcacctccg 2640
tgccgcccac agcctatgag ggaccaagca cctccgtggt gcccaccect gatgagggac 2700
caagcacctc cgtgctgcct acacctggtg agggaccagg cacctccgtg ccgctcgccg 2760
ccactgaggg cctgagcacc tccgtgcagg ccactcctga tgagggaccg agcacctccg 2820
tgccgcccac cgccactgag ggcctaagca ccccgtgcc acccacccgt gatgagggac 2880
cgagcacctc cgtgccggcc actcctggtg agggaccgag cacctccgtg ctgcccgccg 2940
cctctgacgg acaaagcatc tccttggtgc ccaccgcgg taagggatca agcacctccg 3000
tgcccccac cgccaccgag ggcctgagca cctccgtgca gcccactgct ggtgagggat 3060
cgagcacctc cgtgccgccc accctggtg ggactgag cacctccgtg ccgcccaccg 3120
ccactgagga gttgagcacc tccgtgccgc ccactcccgg tgaggcacca agcacttccg 3180
tactgccaat ccccggtgag ggactgagca cctctgtgcc gcccaccgcc tctgatggat 3240
cggcacctc cgtgccgccc actcctggtg agggcgcaag caccttagtg cagcccaccg 3300
cccctgacgg accgggaagc tccgtgctgc ctaaccctgg tgagggcccg agcacattgt 3360
ttagctctag tgcttctgtg gaccggaacc cctccaagtg ttcccttgtt ttgccaagcc 3420
ctagggtaac caaggcctcc gtggactcag attctgggta tcctaagggt gcagaaggcc 3480
ctatagaatt cgaggtcctg agagactgtg agagccccaa ctccattagt attatgggcc 3540
tcaatacttc ccgggttgca attaccctga gccccaagga ccctatggaa cagaacgtag 3600
ctgagctgtt gcagttcctg ctggtgaagg atcagagcaa gtaccctatc cgggagtctg 3660
aaatgcggga atatattgtt aaagaatatc gcaaccagtt tcctgagata ctcaggcgag 3720
```

```
cagcagccca cctggagtgc attttttaggt ttgaattgag agaacttgac cctgaggcac 3780
acacctacat tctgttaaac aaactgggac ctgtgccctt tgaagggtta gaagagagcc 3840
caaatgggcc aaagatgggc ctcctgatga tgattctagg ccaaatattc ctgaatggca 3900
accaagccaa ggaggctgag atttgggaaa tgctctggag gatggggggtg cagcgggaaa 3960
ggaggctttc catttttggg aacccaaaga gacttctgtc tgtggagttt gtatggcagc 4020
gttacttaga ctacaggcca gtaactgact gtaaaccagt ggagtatgag tttttctgggg 4080
gcccaagatc ccacctagaa accaccaaga tgaaaattct gaagttcatg gcgaaaatat 4140
ataacaaaga tcctatggat tggccagaga aatacaacga agctctggaa gaagatgctg 4200
ccagagcctt tgctgagggt tggcaggctc tccctcactt taggaggccc tttttttgagg 4260
aagctgctgc agaggtacca tcccctgatt cagaggtttc cagctattcc tcaaaatatg 4320
ccccacattc atggcctgag tcaagattgg agagcaaggc aaggaagctg gtgcagttat 4380
ttctgcttat ggattcaact aagctgccta taccaaagaa aggaattctg tactacattg 4440
gccgagagtg cagcaaagtg ttccctgacc tcctgaatcg tgctgcccgc accctgaacc 4500
atgtctatgg gacagaacta gtggtacttg atcccaggaa tcactcctat actctgtaca 4560
accgaagggga gatggaagaa actgaggaga tcgtagacag tccaaacagg cctggcaaca 4620
actttttgat gcaggtccta agcttcatct ttattatggg caaccatgcc agggagtctg 4680
cagtctgggc ctttctgcgg ggcttagggg ttcaagctgg gagaaagcat gtgattacct 4740
gcagatactt gagtcagcgc tatatagaca gtttacgggt tcctgacagt gatccagtgc 4800
aatatgagtt tgtatgggggt cctagagccc gtttggaaac ctctaagatg aaagccttgc 4860
gatatgtggc cagaatccac agaaaggaac cacaggactg gccacagcag tacagggagg 4920
caatggaaga tgaggccaat agagctgatg ttgggcacag gcaaatcttt gttcacaact 4980
tcaggtagag gaatgcatgg cagtcagagg ggccttgcaa ggaggggcct ttgagcctca 5040
gttctcatgt attggggggggt ggggggtgggt acatattgta tttggtattt gtgttccagt 5100
tatatttatg tcttttttcata tttagttctg gtgtggtgtc tggaaatgtt tcaactgttt 5160
taatatattg tctgattggg taaatgtgat tgaccacttg ctgtctctgt tgtattttgg 5220
tatgagtttt gatagctcta taaaatgttt tggaaatctt tccatcttgt tgcattatct 5280
agaaaagaat atagcatata gctatagata taggcttttc cttgaaagct tgaaaaaaat 5340
tcgccagtaa aataatctag cttaaagtga taaactaaca aacaaatata acaacaactg 5400
aaaaggcact ctaaattgtg gctggcttcc ctttctgtct gctattgtat gaagaatact 5460
gatgtttacc tgtatttgct ttgctgtact aaaatgtaat gaaaaataaa agattaataa 5520
atgaaacata atgctaata                                             5539
```

<210> 46.
<211> 2338
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6803363CB1

<400> 46
```
cggaggggtg ctaacggtgc taacgctgcc cgctccccgc ccgccgcag gtgcgcagga 60
ggatggtggc gcggccctag gcccacgctc cgcaccatga cctgctggct gtgcgtcctg 120
agcctgcccc tgctcctgct gcccgcggcg ccgcccccgg ctggaggctg cccggcccgc 180
tgcgagtgca ccgtgcagac ccgcgcggtg gcctgcacgc gccgccgcct gaccgccgtg 240
cccgacggca tcccggccga gaccgcctg ctggagctca gccgcaaccg catccgctgc 300
ctgaacccgg gcgacctggc cgcgctgccc gcgctggagg agctggacct gagcgagaac 360
gccatcgcgc acgtggagcc cggcgccttc gccaacctgc cgcgcctgcg cgtcctgcgt 420
ctccgtggca accagctgaa gctcatcccg cccggggtct tcacgcgccct ggacaacctc 480
acgctgctgg acctgagcga gaacaagctg gtaatcctgc tggactacac tttccaggac 540
ctgcacagcc tgcgccggct ggaagtgggc gacaacgacc tggtattcgt ctcgcgccgc 600
gccttcgcgg ggctgctggc cctggaggag ctgaccctgg agcgctgcaa cctcacggct 660
ctgtccggggg agtcgctggg ccatctgcgc agcctgggcg ccctgcggct cgccacctg 720
gccatcgcct ccctggagga ccagaacttc cgcaggctgc ccgggctgct gcacctggag 780
attgacaact ggccgctgct ggaggaggtg gcggcgggca gcctgcgggg cctgaacctg 840
acctcgctgt cggtcaccca caccaacatc accgccgtgc cggccgccgc gctgcggcac 900
caggcgcacc tcacctgcct caatctgtcg cacaacccca tcagcacggt gccgcggggg 960
tcgttccggg acctggtccg cctgcgcgag ctgcacctgg ccgggggccct gctggctgtg 1020
gtggagccgc aggccttcct gggcctgcgc cagatccgc tgctcaacct ctccaacaac 1080
ctgctctcca cgttggagga gagcaccttc cactcggtga acacgctaga gacgctgcgc 1140
gtggacggga acccgctggc ctgcgactgt cgcctgctgt ggatcgtgca gcgtcgcaag 1200
accctcaact tcgacgggcg gctgccggcc tgcgccaccc cggccgaggt gcgcggcgac 1260
```

```
gcgctgcgaa  acctgccgga  ctccgtgctg  ttcgagtact  tcgtgtgccg  caaacccaag  1320
atccgggagc  ggcggctgca  gcgcgtcacg  gccaccgcgg  gcgaagacgt  ccgcttcctc  1380
tgccgcgccg  agggcgagcc  ggcgcccacc  gtggcctggg  tgacccccca  gcaccggccg  1440
gtgacggcca  ccagcgcggg  ccgggcgcgc  gtgctccccg  ggggacgct  ggagatccag  1500
gacgcgcggc  cgcaggacag  cggcacctac  acgtgcgtgg  ccagcaacgc  gggcggcaac  1560
gacacctact  tcgccacgct  gaccgtgcgc  cccgagccgg  ccgccaaccg  gacccggggc  1620
gaggcccaca  acgagacgct  ggcggccctg  cgcgcgccac  tcgacctcac  caccatcctg  1680
gtgtccaccg  ccatgggctg  catcaccttc  ctgggcgtgg  tcctcttctg  cttcgtgctg  1740
ctgttcgtgt  ggagccgcgg  ccgcgggcag  cacaaaaaca  acttctcggg  ggagtactcc  1800
ttccgcaagg  tggatgggcc  ggccgccgcg  gcgggccagg  gaggcgcgcg  caagttcaac  1860
atgaagatga  tctgaggggt  ccccagggcg  gaccctcccc  tcccctcccc  cgcgggccgg  1920
ccgctcgcgt  gtccacctat  gcatttcccg  gaggggaagg  ggacggctgc  acggcgctcc  1980
ccaggcagaa  cttcccctt  ttttgtagac  gcccaaccgc  aggactgttt  ttcatcagca  2040
tgcgtctttt  ttgcagtttc  tcaagcgttt  tctaaagatt  tcaacccgtc  ttcctgtccc  2100
tgctatggtg  gctgagctgg  ggggtggggg  tgcgggctgc  aaggatgggg  ggatgggctc  2160
ccgtctacca  acccgggggtg  gctggggggcg  ctcagagacc  ccaggatcct  caggggccgc  2220
aattcacacc  caggaggcac  aagactgctg  cccccaagcc  ctgggcttcc  gctcccccag  2280
ccctttgtac  acagatcgca  tcggggagac  ggtggacacc  cgggtcccac  aaagggcc    2338
```

```
<210> 47
<211> 3214
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7070580CB1

<400> 47
ggagatgccc  cttcaaggct  tttgcccttc  ctgagggagc  agaaaaagcc  acgaagactc  60
ctcgtaaaca  acttgagaga  gagaattgtg  tctccctat  gaagatggat  tgcttgagac  120
agagttcaat  agaagatgtt  aaacagataa  gctgagggca  gggggactga  gctttggaag  180
agggcctttc  aggcagcgag  gccgctggcc  tgtggctccc  tctgatgagc  ttcttctaaa  240
gatggtaagg  gctggggggaa  gagcttggca  gcaggggctg  cactctctct  cagctgtgtc  300
tatcttggcc  aagggtttta  tgttcctatt  ggtaggggat  tcgagagcag  agagcgtcac  360
caatacactc  gtgttgttca  ccataggaga  agagtccctg  accattttg  tggacaagca  420
gaaactggga  agaaagacag  agacaacagg  aggtgcctct  ataatcgggg  gcagtgggaa  480
cagcacagct  gtgtccctgg  agaccctgca  ccagctggcc  gcctcctact  tcatcgacag  540
agagagcacg  ctgcgacggc  tgcaccatat  ccagatagcc  acgggggcca  tcaaggtcac  600
cgagaccagg  accggtcctc  tgggctgcag  caactatgac  aatctggact  cagtcagttc  660
tgtcttggta  cagagtccag  agaacaaagt  acagttactt  ggccttcagg  tgctgctgcc  720
tgagtatctg  cgtgagcgct  ttgtagctgc  tacatcacat  gcagctctga  780
gggtgagctc  gtctgcaagg  agaatgactg  ctggtgcaag  tgcagcccca  ccttccctga  840
atgcaactgc  cctgatgctg  acatccaggc  catggaggac  agcctgctgc  agatccagga  900
ctcctgggcc  actcacaacc  ggcagtttga  agagtcagaa  gagttccagg  ccctgctgaa  960
aaggctgccc  gatgaccggt  tcctgaactc  cacagctatc  tcccagttct  gggccatgga  1020
caccagcctt  cagcaccgct  accagcagct  gggagctggc  ttgaaagtgc  tgttcaaaaa  1080
gacccatcgg  atcctacgcc  ggctcttcaa  cctctgcaag  cgctgccatc  gccagcctcg  1140
cttccgcctg  cccaaggaga  ggtccttgtc  ctactggtgg  aaccgaatcc  agtccctcct  1200
ctactgtggg  gaaagcacct  ttcctggcac  tttcctggaa  cagagccaca  gctgcacctg  1260
cccctatgac  caatcttcct  gccagggccc  catcccatgt  gccttgggcg  aagggcccgc  1320
gtgtgcccac  tgtgctccag  acaatagcac  acgctgtggg  agctgcaacc  cgggctatgt  1380
gctggcccag  gggctgtgcc  ggccagaggt  ggccgagtcc  ctggaaaact  ttcttgggct  1440
ggagacagac  ttgcaggacc  tggagctaaa  gtacctgctg  cagaagcagg  atagccgcat  1500
tgaggtacac  tccatcttca  tcagcaatga  catgcggctg  ggcagctggt  ttgacccttc  1560
ctggaggaag  cgcatgctgc  tcaccctgaa  gagcaacaag  tacaagcctg  ggctggtgca  1620
cgtgatgttg  gccttgtcct  tgcagatctg  tctcaccaag  aacagcaccc  tggagcctgt  1680
catggccatc  tacgtcaacc  ccttggggg  cagccactct  gagagctggt  tcatgcctgt  1740
gaatgagggc  agctttcctg  actgggaaag  gactaacgtg  gatgcagctg  cccagtgcca  1800
aaactggact  atcaccttgg  ggaataggtg  gaagactttc  tttgagacag  ttcatgttta  1860
cctacggagc  cgaatcaagt  ccctggatga  cagctccaat  gagacaatct  actatgagcc  1920
cctggagatg  actgatccct  ctaagaattt  gggttacatg  aaaattaaca  ccttgcaggt  1980
ctttggctac  agcctgccct  ttgacccaga  tgctatccgg  gacttaattc  tccagttcga  2040
```

```
ctacccatat actcaaggtt cccaggactc tgcactcttg cagctcattg agctcaggga 2100
ccgggtgaac cagctttctc cacctggcaa agtccgactt gaccttttct cctgcttgct 2160
ccggcatcgg cttaagctgg ccaacaatga ggtgggcagg atccagtcct ccctgagggc 2220
tttcaattct aagctgccaa accctgtgga atatgagacc ggcaaactct gtagctaatg 2280
ggcggcccac ttcagcactg ggcaaggagg ggatccatga atctggggta caaagataat 2340
ctaagccctc accttagtgc caacagggtg tgctcccacg agactttcag catccagtag 2400
atgggacctc gaggctcgag ctgaagcagg cgagagagaa acagctactg cgtgcgtgcg 2460
cgcacgcata cacacacaca cacacacact ggcacaggga ggctacaact aagcagcctc 2520
agatctgtaa agttgcattgg tgctttctaa aatgaatgca attgaaagaa aggagccaag 2580
gaagagatta agaaaaagaa ccagctgaac catgaaacaa acaaaaatcc taaaagtgat 2640
tcttgtcatt caagaaagca agaggggaca acaatgggag gggtcggagc tcttccttcc 2700
cctctgtgga gtcacttttg tattcttttt aaccagattt cttaaaatgt tgttgttttg 2760
tgaatcctga cattggttct tacttttgta tgctgcctcc tctgtgccct cccagacgct 2820
gactgggaaa cacaagaagt acaaccaaca ggaaccagcg ccaagggcag gcagcggcct 2880
ccttgctccc ctcccttact cctccctctg ctgcctcctc cccccaccaa gtttcagggc 2940
cctggattgt tcccagttcc cattgtggtc ccttcagagc tcctttccaa cagcatctct 3000
ctgtcgaaga aagaagctct gtcaagttag agagagacaa tgtgtaggaa atgttctttt 3060
ttaaaaaaaa ataacaaaaa caaaacaaaa ctatttattt tgtgattgtt ttccttgtta 3120
atctgctcca accacctgaa catctaagta aacatttatc tggtttaata taaaaaaaaa 3180
aaaaaaaaaa attctgggcc caagaattag ctgt                          3214
```

<210> 48
<211> 2824
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500176CB1

<400> 48
```
cagcgattga tatttctaat cgtcgtgatc cacctccccca ggaccttgga gccacgttta 60
caaataggaa tagggtacgt gggagggata gaacgtacag ccaataaaat catgtggcgc 120
cgatgggcgt gttgaggccg ctgcctggct tagggcggaa acagattctc tgcataagaa 180
ggggaacgaa agatggcggc ggaaacgctg ctgtccagtt tgttaggact gctgcttctg 240
ggactcctgt tacccgcaag tctgaccggc ggtgtcggga gcctgaacct ggaggagctg 300
agtgagatgc gttatgggat cgagatcctg ccgttgcctg tcatgggagg gcagagccaa 360
tcttcggacg tggtgattgt ctcctctaag tacaaacagc gctatgagtg tcgcctgcca 420
gctggagcta ttcacttcca gcgtgaaagg gaggaggaaa cacctgctta ccaagggcct 480
gggatccctg agttgttgag cccaatgaga gatgctccct gcttgctgaa gacaaaggac 540
tggtggacat atgaattctg ttatggacgc cacatccagc aataccacat ggaagattca 600
gagatcaaag gtgaagtcct ctatctcggc tactaccaat cagccttcga ctgggatgat 660
gaaacagcca aggcctccaa gcagcatcgt cttaaacgct accacagcca gacctatggc 720
aatgggtcca agtgcgacct taatgggagg ccccgggagg ccgaggttcg gttcctctgt 780
gacgagggtg caggtatctc tggggactac atcgatcgcg tggacgagcc cttgtcctgc 840
tcttatgtgc tgaccattcg cactcctcgg ctctgccccc accctctcct ccggcccccca 900
cccagtgctg caccgcaggc catcctctgt cacccttccc tacagcctga ggagtacatg 960
gcctacgttc agaggcaagc cgactcaaag cagtatggag ataaaatcat agaggagctg 1020
caagatctag gcccccaagt gtggagtgag accaagtctg gggtggcacc ccaaaagatg 1080
gcaggtgcga gcccgaccaa ggatgacagt aaggactcag atttctggaa gatgcttaat 1140
gagccagagg accaggcccc aggaggggag gaggtgccgg ctgaggagca ggacccaagc 1200
cctgaggcag cagattcagc ttctggtgct cccaatgatt ttcagaacaa cgtgcaggtc 1260
aaagtcattc gaagccctgc ggatttgatt cgattcatag aggagctgaa aggtggaaca 1320
aaaaagggga agccaaatat aggccaagag cagcctgtgg atgatgctgc agaagtccct 1380
cagagggaac cagagaagga aaggggtgat ccagaacggc agagagagat ggaagaagag 1440
gaggatgagg atgaggatga ggatgaagat gaggatgaac ggcagttact gggagaattt 1500
gagaaggaac tggaagggat cctgcttccg tcagaccgag accggctccg ttcggagaca 1560
gagaaagagc tggacccaga tgggctgaag aaggagtcag agcgggatcg ggcaatgctg 1620
gctctcacat ccactctcaa caaactcatc aaaagactgg aggaaaaaca gagtccagag 1680
ctggtgaaga agcacaagaa aaagagggtt gtcccaaaaa agcctccccc atcaccccaa 1740
cctacagagg aggatcctga gcacagagtc cgggtccggg tcaccaagct ccgtctcgga 1800
ggccctaatc aggatctgac tgtcctcgag atgaaacggg aaaacccaca gctgaaacaa 1860
atcgagggggc tggtgaagga gctgctggag agggagggac tcacagctgc agggaaaatt 1920
```
```

```
gagatcaaaa ttgtccgccc atgggctgaa gggactgaag agggtgcacg ttggctgact 1980
gatgaggaca cgagaaacct caaggagatc ttcttcaata tcttggtgcc gggagctgaa 2040
gaggcccaga aggaacgcca gcggcagaaa gagctggaga gcaattaccg ccggggtgtgg 2100
ggctctccag gtggggaggg cacaggggac ctggacgaat ttgacttctg agaccaacac 2160
tacacttgac ccttcacgga atccagactc ttcctggact ggcttgcctc ctccccacct 2220
ccccaccctg gaacccctga gggccaaaca gcagagtgga gctgagctgt ggacctctcg 2280
ggcaactctg tgggtgtggg ggccctgggt gaatgctgct gcccctgctg gcagccacct 2340
tgagacctca ccgggcctgt gatatttgct ctcctgaact ctcactcaat cctcttcctc 2400
tcctctgtgg cttttcctgt tattgtcccc taatgatagg atattccctg ctgcctacct 2460
ggagattcag taggatcttt tgagtggagg tgggtagaga gagcaaggag ggcaggacac 2520
ttagcaggca ctgagcaagc aggcccccac ctgcccttag tgatgtttgg agtcgtttta 2580
ccctcttcta ttgaattgcc ttgggattttc cttctccctt tccctgccca ccctgtcccc 2640
tacaatttgt gcttctgagt tgaggagcct tcacctctgt tgctgaggaa atggtagaat 2700
gctgcctatc acctccagca caatcccagt gaaaaaggtg tgaagcaccc accatgttct 2760
tgaacaatca ggtttctaaa taaacaactg gaccatcaaa gaaaagaaaa aaaaaacgcg 2820
gcgg                                                             2824
```

<210> 49
<211> 1570
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500506CB1

<400> 49

```
ctcagtggct ggagggcagg gccagggtgc ggacacaagc tgcagacaca gaccatggcc 60
ttgccaacgg ctcgacccct gttggggtcc tgtgggaccc ccgccctcgg cagcctcctg 120
ttcctgctct tcagcctcgg atgggtgcag ccctcgagga ccctggctgg agagacaggg 180
caggaggctg cgccctggga cggagtcctg gccaacccac ctaacatttc cagcctctcc 240
cctcgccaac tccttggctt cccgtgtgcg gaggtgtccg gcctgagcac ggagcgtgtc 300
cgggagctgg ctgtggcctt ggcacagaag aatgtcaagc tctcaacaga gcagggcatc 360
gtggccgcgt ggcggcaacg ctcctctcgg gacccatcct ggcggcagcc tgaacggacc 420
atcctccggc cgcggttccg gcgggaagtg gagaagacag·cctgtccttc aggcaagaag 480
gcccgcgaga tagacgagag cctcatcttc tacaagaagt gggagctgga gcctgcgtg 540
gatcggccc tgctggccac ccagatggac cgcgtgaacg ccatcccctt cacctacgag 600
cagctggacg tcctaaagca taaactggat gagctctacc acaaggtta ccccgagtct 660
gtgatccagc acctgggcta cctcttcctc aagatgagcc ctgaggacat tcgcaagtgg 720
aatgtgacgt ccctgaaggac cctgaaggct ttgcttgaag tcaacaaagg gcacgaaatg 780
agtcctcagg tggccaccct gatcgaccgc tttgtgaagg gaagggggcca gctagacaaa 840
gacaccctag acaccctgac cgccttctac cctgggtacc tgtgctccct cagcccccgag 900
gagctgagct ccgtgccccc cagcagcatc tgggcggtca ggccccagga cctggacacg 960
tgtgacccaa ggcagctgga cgtcctctat cccaaggccc gccttgcttt ccagaacatg 1020
aacgggtccg aatacttcgt gaagatccag tccttcctgg gtggggcccc cacggaggat 1080
ttgaaggcgc tcagtcagca gaatgtgagc atggacttgg ccacgttcat gaagctgcgg 1140
acggatgcgg tgctgccgtt gactgtggct gaggtgcaga aacttctggg accccacgtg 1200
gagggcctga aggcggagga gcggcaccgc ccggtgcggg actggatcct acggcagcgg 1260
caggacgacc tggacacgct ggggctgggg ctacagggcg gcatccccaa cggctacctg 1320
gtcctagacc tcagcgtgca agaggccctc tcggggacgc cctgcctcct aggacctgga 1380
cctgttctca ccgtcctggc actgctccta gcctccaccc tggcctgagg gccccactcc 1440
cttgctggcc ccagccctgc tggggatccc cgcctggcca ggagcaggca cgggtgatcc 1500
ccgttccacc ccaagagaac tcgcgctcag taaacgggaa catgccccct gcagacaaaa 1560
aaaaaaaaaa                                                       1570
```

<210> 50
<211> 725
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500639CB1

<400> 50
ctggagaggt ggggagctgc tgaagggcgg gagaagttgc ggcgtggcga taggtctggg 60
tgttctggga ggacccctaa aagggagcag ctctgccagg tcagagcttg aggtagcatg 120
tgtgccgctt ccaggaggcc aggatggaga cagcagttcg aggaatgccc ttggaatgcc 180
ctcctaggcc ggagcggctc aatgcctacg agcgcgaagt gatggtgaac atgctgaact 240
cactgtcgcg gaaccagcag ctgccgcgga tcacgccccg atgcgggtgc gtggacccgc 300
tgcccggccg cctgccccttc catggttacg aaagtgcttg ctcgggccgc cactactgtc 360
tgcgcgggat ggactactac gccagcgggg cgccctgcac cgaccgccgg ctgcggcctt 420
ggtgccggga gcaaccgact atgtgtacct ccctacgagc accggcccgg aatgcagtgt 480
gctgttacaa ctcccccgcc gtcatactac ccatatccga accttagatg ggacacaagt 540
cacttcaaga agtctggtgg tccccagaga aacaactatg ttatccatcc tgagtttgtg 600
tctgagacct atcccgacta tcgttgctgg tagagcctgg gcttgctgtg ccaaggggc 660
agagcaataa ataaactctt caaccccaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 720
aaatg   725

<210> 51
<211> 3108
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506167CB1

<400> 51
gcaagagtgt ctgtgtgcat atcagataaa gcaaagaaaa ctaccaaaac cagagattat 60
ataatgatca aaggtttaat ccatgaagaa gacagcaatt ttaaaagtgt attcaccaaa 120
aaataaagct tcaaaatatg tgatgtgaaa actgccagaa ctaaggcggg ccgggctcag 180
accagcgctg cctcaggatg tgaagtgtaa·caagagggcc aggggaggtg gtggggggaca 240
acatgggcct gtgaggcctg tgggtgcccg cgttccccag ctcccccccgc agcccgctcc 300
acagtggtcc gctccggttg gttgtcacgt gcgcattcgg gttccagacc caaggctgcg 360
tgttctccac cgcttgttgt ggccagtgtt actgcggtga ccgccagagc agcctcgacg 420
ctatggagga gcctggtgct accccctcagc cctacctggg gctggtcctg gaggagctac 480
gcagagttgt ggcagcacta·cctgagagta tgagaccaga tgagaatcct·tatggttttc 540
catcggaact ggtggtatgt gcagctgtta ttggattttt tgttgttctc cttttttttgt 600
ggagaagttt tagatcggtt aggagtcggc tttacgtggg aagagagcaa aaacttggtg 660
caacgctttc tggactaatt gaagaaaaat gtaaactact tgaaaagttt agccttattc 720
aaaaagagta tgaaggctat gaagtagagt catctttaga ggatgccagc tttgagaagg 780
cggcagcaga agaagcacga agtttggagg caacctgtga aaagctgaac aggtccaatt 840
ctgaacttga ggatgaaatc ctctgtctag aaaaagactt aaaacaagag aaatctaaac 900
attctcaaca agatgaattg atggcggata tttcaaaaag tatacagtct ctagaagatg 960
agtcaaaatc cctcaaatca caaatagctg aagccaaaat catctgcaag acatttaaaa 1020
tgagtgaaga acgacgggct atagcaataa aagatgcttt gaatgaaaat tctcaacttc 1080
agacaagcca taaacagctt tttcagcaag aagctgaagt atggaaagga gaagtgagtg 1140
aacttaataa acagaaaata·acatttgaag actccaaagt acacgcagaa caagttctga 1200
atgataaaga aaatcacatc aagaccctga ctggacactt gccaatgatg aaagatcagg 1260
ctgctgtgct tgaagaagac acaacggatg atgataacct ggaattagaa gtgaacagtc 1320
aatgggaaaa tggtgcttac ttagataatc ctccaaaagg agctttgaag aaactgattc 1380
atgctgctaa gttaaatgct tctttaaaaa ccttagaagg agaaagaaac caaatttata 1440
ttcagttgtc tgaagttgat aaaacaaagg aagagcttac agagcatatt aaaaatcttc 1500
agactcaaca agcatctttg cagtcagaaa acacacattt tgaaaatgag aatcagaagc 1560
ttcaacagaa acttaaagta atgactgaat tatatcaaga aaatgaaatg aaactccaca 1620
ggaaattaac agtagaggaa aattatcggt tagagaaaga agagaaactt tctaaagtag 1680
atgaaaagat cagccatgcc actgaagagc tggagaccta tagaaagcga gccaaagatc 1740
ttgaagaaga attggagaga actattcatt cttatcaagg gcagattatt tcccatgaga 1800
aaaaagcaca tgataattgg ttggcagctc ggaatgctga agaaacctc aatgatttaa 1860
ggaaagaaaa tgctcacaac agacaaaaat taactgaaac agagcttaaa tttgaacttt 1920
tagaaaaaga tccttatgca ctcgatgttc caaatacagc atttggcaga ggctcacgag 1980
gcccagggaa tcctctggac catcagatta ccaatgaaag aggagaatca agctgtgata 2040
ggttaaccga tcctcatagg gctccctctg acactggtc tctgtcacct ccatgggacc 2100
aggaccgtag gatgatgttt cctccgccag gacaatcata tcctgattca gcccttcctc 2160
cacaaaggca agacagattt tgttctaatt ctggtagact gtctggacca gcagaactca 2220
gaagttttaa tatgccttct ttggataaaa tggatgggtc aatgccttca gaaatggaat 2280

68/71

```
ccagtagaaa tgataccaaa gatgatcttg gtaatttaaa tgtgcctgat tcatctctcc 2340
ctgctgaaaa tgaagccact ggccctggct ttgttcctcc acctcttgct ccaatcagag 2400
gtccattgtt tccagtggat gcaagaggcc cattcttgag aagaggacct cctttccccc 2460
cacctcctcc aggagccatg tttggagctt ctcgagatta tttccacca agggatttcc 2520
caggtccacc acctgctcca tttgcaatga gaaatgtcta tccaccgagg ggttttcctc 2580
cttaccttcc cccaagacct ggattttttcc ccccacccc acattctgaa ggtagaagtg 2640
agttcccctc aggtttgatt ccaccttcaa atgagcctgc tactgaacat ccagaaccac 2700
agcaagaaac ctgacaatat tttgctctc ttcaaaagta attttgactg atctcatttt 2760
cagtttaagt aactgctgtt acttaagtga ttacacttttt gctcaaattg aagcttaatg 2820
gaattataat tctcaggata gtattttgta aataaagatg atttaaatat gaatcttatg 2880
agtaaattat ttcaattta ttttagacgg tataactatt tcaatttgat taatccacta 2940
ttatataaac aatagtggga gttttatata tgtaatcttt caggtggggga ggctttaaat 3000
tctgaagtct gtgtctttat gccaagaact gtatttactg tggttgtgga caaatgtgaa 3060
agtaacttta tgcttaaata aattatagtt gatttaaaga tttgtttg          3108
```

<210> 52
<211> 1962
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 90081189CB1

<400> 52

```
gccgctgcct ggcttagggc ggaaacagat tctctgcata agaaggggaa cgaaagatgg 60
cggcggaaac gctgctgtcc agtttgttag gactgctgct tctgggactc ctgttacccg 120
caagtctgac cggcggtgtc gggagcctga acctggagga gctgagtgag atgcgttatg 180
ggatcgagat cctgccgttg cctgtcatgg gagggcagag ccaatcttcg gacgtggtga 240
ttgtctcctc taagtacaaa cagcgctatg agtgtcgcct gccagctgga gctattcact 300
tccagcgtga aagggaggag gaaacacctg cttaccaagg gcctgggatc cctgagttgt 360
tgagcccaat gagagatgct ccctgcttgc tgaagacaaa ggactggtgg acatatgaat 420
tctgttatgg acgccacatc cagcaatacc acatggaaga ttcagagatc aaaggtgaag 480
tcctctatct cggctactac caatcagcct tcgactggga tgatgaaaca gccaaggcct 540
ccaagcagca tcgtcttaaa cgctaccaca gccagaccta tggcaatggg tccaagtgcg 600
acctaatgg gaggccccgg gaggccgagg ttcggttcct ctgtgacgag ggtgcaggta 660
tctctgggga ctacatcgat cgcgtggacg agcccttgtc ctgctcttat gtgctgacca 720
ttcgcactcc tcggctctgc ccccacccctc tcctccggcc cccacccagt gctgcaccgc 780
aggccatcct ctgtcaccct tccctacagc ctgaggagta catggcctac gttcagaggc 840
aagccgactc aaagcagtat ggagataaaa tcatagagga gctgcaagat ctaggccccc 900
aagtgtggag tgagaccaag tctggggtgg caccccaaaa gatggcaggt gcgagcccga 960
ccaaggatga cagtaaggac ccagatttct ggaagatgct taatgagcca gaggaccagg 1020
ccccaggagg ggaggaggtg ccggctgagg agcagtgacc aagccctgag gcagcagatt 1080
cagcttctgg tgctcccaat gattttcaga acaacgtgca ggtcaaagtc attcgaagcc 1140
ctgcggattt gattcgattc atagaggagc tgaaaggtgg aacaaaaaag gggaagccaa 1200
atataggcca agagcagcct gtggatgatg ctgcagaagt ccctcagagg gaaccagaga 1260
aggaaagggg tgatccagaa cggcagagag agatggaaga agaggaggat gaggatgagg 1320
atgaggatga agatgaggat gaacggcagt tactgggaga atttgagaag gaactggaag 1380
ggatcctgct tccgtcagac cgagaccggc tccgttcgga ggtgaaggct ggcatggagc 1440
gggaactgga aaacatcatc caggagacag agaaagagct ggaccagat gggctgaaga 1500
aggagtcaga gcgggatcgg gcaatgctgg ctctcacatc cactctcaac aaactcatca 1560
aaagactgga ggaaaacag agtccagagc tggtgaagaa gcacaagaaa aagaggggttg 1620
tccccaaaaa gcctcccca tcaccccaac ctacagggaa aattgagatc aaaattgtcc 1680
gcccatgggc tgaagggact gaagagggtg cacgttggct gactgatgag gacacgagaa 1740
acctcaagga gatcttcttc aatatcttgg tgccgggagc tgaagaggcc cagaaggaac 1800
gccagcggca gaaagagctg gagagcaatt accgccgggt gtggggctct ccaggtgggg 1860
agggcacagg ggacctggac gaatttgact tctgagacca acactacact tgacccttca 1920
cggaatccag actcttcctg gactggcttg cctcctcccc ac             1962
```

<210> 53
<211> 3005
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7510095CB1

<220>
<221> unsure
<222> 2808
<223> a, t, c, g, or other

<400> 53
```
cagcgattga tatttctaat cgtcgtgatc cacctcccca ggaccttgga gccacgttta 60
caaataggaa tagggtacgt gggagggata gaacgtacag ccaataaaat catgtggcgc 120
cgatgggcgt gttgaggccg ctgcctggct tagggcggaa acagattctc tgcataagaa 180
ggggaacgaa agatggcggc ggaaacgctg ctgtccagtt tgttaggact gctgcttctg 240
ggactcctgt tacccgcaag tctgaccggc ggtgtcggga gcctgaacct ggaggagctg 300
agtgagatgc gttatgggat cgagatcctg ccgttgcctg tcatgggagg gcagagccaa 360
tcttcggacg tggtgattgt ctcctctaag tacaaacagc gctatgagtg tcgcctgcca 420
gctggagcta ttcacttcca gcgtgaaagg gaggaggaaa cacctgctta ccaagggcct 480
gggatccctg agttgttgag cccaatgaga gatgctccct gcttgctgaa gacaaaggac 540
tggtggacat atgaattctg ttatggacgc cacatccagc aataccacat ggaagattca 600
gagatcaaag gtgaagtcct ctatctcggc tactaccaat cagccttcga ctgggatgat 660
gaaacagcca aggcctccaa gcagcatcgt cttaaacgct accacagcca gacctatggc 720
aatgggtcca agtgcgacct taatgggagg ccccgggagg ccgaggttcg gttcctctgt 780
gacgagggtg caggtatctc tggggactac atcgatcgcg tggacgagcc cttgtcctgc 840
tcttatgtgc tgaccattcg cactcctcgg ctctgccccc accctctcct ccggccccca 900
cccagtgctg caccgcaggc catcctctgt cacccttccc tacagcctga ggagtacatg 960
gcctacgttc agaggcaagc cgactcaaag cagtatggag ataaaatcat agaggagctg 1020
caagatctag gcccccaagt gtggagtgag accaagtctg gggtggcacc ccaaaagatg 1080
gcaggtgcga gcccgaccaa ggatgacagt aaggactcag atttctggaa gatgcttaat 1140
gagccagagg accaggcccc aggaggggag gaggtgccgg ctgaggagca ggacccaagc 1200
cctgaggcag cagattcagc ttctggtgct cccaatgatt ttcagaacaa cgtgcaggtc 1260
aaagtcattc gaagccctgc ggatttgatt cgattcatag aggagctgaa aggtggaaca 1320
aaaaagggga agccaaatat aggccaagag cagcctgtgg atgatgctgc agaagtccct 1380
cagagggaac cagagaagga aaggggtgat ccagaacggc agagagagat ggaagaagag 1440
gaggatgagg atgaggatga ggatgaagat gaggatgaac ggcagttact gggagaattt 1500
gagaaggaac tggaagggat cctgcttccg tcagaccgag accggctccg ttcggaggtg 1560
aaggctggca tggagcggga actggaaaac atcatccagg agacagagaa agagctggac 1620
ccagatgggc tgaagaagga gtcagagcgg gatcgggcaa tgctggctct cacatccact 1680
ctcaacaaac tcatcaaaag actggaggaa aaacagagtc cagagctggt gaagaagcac 1740
aagaaaaaga gggttgtccc caaaaagcct cccccatcac cccaacctac agaggaggat 1800
cctgagcaca gagtccgggt ccgggtcacc aagctccgtc tcggaggccc taatcaggat 1860
ctgactgtcc tcgagatgaa acgggaaaac ccacagctga aacaaatcga ggggctggtg 1920
aaggagctgc tggtagaggga gggactcaca gctgcaggtg ggccctggag ggcggctgga 1980
cccagtgctg tcggaagggc aagctgccgg aagtggaggg gctgggacca gtggggcagg 2040
gggcctgccc tgaggctccc aaccatcctc tcctctcact gcatactgct cctttccagg 2100
gaaaattgag atcaaaattg tccgcccatg ggctgaaggg actgaagagg gtgcacgttg 2160
gctgactgat gaggacacga gaaacctcaa ggagatcttc ttcaatatct tggtgccggg 2220
agctgaagag gcccagaagg aacgccagcg gcagaaagag ctggagagca attaccgccg 2280
ggtgtgggggc tctccaggtg gggagggcac aggggacctg gacgaatttg acttctgaga 2340
ccaacactac acttgaccct tcacggaatc cagactcttc ctggactggc ttgcctcctc 2400
cccacctccc caccctggaa cccctgaggg ccaaacagca gagtggagct gagctgtgga 2460
cctctcgggc aactctgtgg gtgtgggggc cctgggtgaa tgctgctgcc cctgctggca 2520
gccaccttga gacctcaccg ggcctgtgat atttgctctc ctgaactctc actcaatcct 2580
cttcctctcc tctgtggctt ttcctgttat tgtcccctaa tgataggata ttccctgctg 2640
cctacctgga gattcagtag gatcttttga gtggaggtgg gtagagagag caggagggc 2700
aggacactta gcaggcactg agcaagcagg cccccacctg cccttagtga tgtttggagt 2760
cgttttaccc tcttctattg aattgccttg ggatttcctt ctccttntc ctgcccaccc 2820
tgtcccctac aatttgtgct tctgagttga ggagccttca actctgttgc tgaggaaatg 2880
gtagaatgct gcctatcaac ctcagcaaaa atccagcgaa aaaggtgtgt aagcaaccac 2940
caatgtctct ggaacaaatc agggtttaa ataaacacct tgggcccatc aaaaaaaaaa 3000
aaaaa                                                         3005
```

<210> 54

<211> 2906
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7510096CB1

<400> 54
```
cagcgattga tatttctaat cgtcgtgatc cacctcccca ggaccttgga gccacgttta 60
caaataggaa tagggtacgt gggagggata gaacgtacag ccaataaaat catgtggcgc 120
cgatgggcgt gttgaggccg ctgcctggct tagggcggaa acagattctc tgcataagaa 180
ggggaacgaa agatggcggc ggaaacgctg ctgtccagtt tgttaggact gctgcttctg 240
ggactcctgt tacccgcaag tctgaccggc ggtgtcggga gcctgaacct ggaggagctg 300
agtgagatgc gttatgggat cgagatcctg ccgttgcctg tcatgggagg gcagcctttc 360
tttgcttcct tgcccccccga ccctcccttc tttcccagag ccaatcttcg gacgtggtga 420
ttgtctcctc taagtacaaa cagcgctatg agtgtcgcct gccagctgga gctattcact 480
tccagcgtga aagggaggag gaaacacctg cttaccaagg gcctgggatc cctgagttgt 540
tgagcccaat gagagatgct ccctgcttgc tgaagacaaa ggactggtgg acatatgaat 600
tctgttatgg acgccacatc cagcaatacc acatggaaga ttcagagatc aaaggtgaag 660
tcctctatct cggctactac caatcagcct cgactgggga tgatgaaaca gccaaggcct 720
ccaagcagca tcgtcttaaa cgctaccaca gccagaccta tggcaatggg tccaagtgcg 780
accttaatgg gaggccccgg gaggccgagg ttcggttcct ctgtgacgag ggtgcaggta 840
tctctgggga ctacatcgat cgcgtggacg agcccttgtc ctgctcttat gtgctgacca 900
ttcgcactcc tcggctctgc ccccaccctc tcctccggcc cccacccagt gctgcaccgc 960
aggccatcct ctgtcaccct tccctacagc ctgaggagta catggcctac gttcagaggc 1020
aagccgtaga ctcaaagcag tatggagata aaatcataga ggagctgcaa gatctaggcc 1080
cccaagtgtg gagtgagacc aagtctgggg tggcacccca aaagatggca ggtgcgagcc 1140
cgaccaagga tgacagtaag gactcagatt tctggaagat gcttaatgag ccagaggacc 1200
aggccccagg aggggaggag gtgccggctg aggagcagga cccaagccct gaggcagcag 1260
attcagcttc tggtgctccc aatgattttc agaacaacgt gcaggtcaaa gtcattcgaa 1320
gccctgcgga tttgattcga ttcatagagg agctgaaagg tggaacaaaa aaggggaagc 1380
caaatatagg ccaagagcag cctgtggatg atgctgcaga agtccctcag agggaaccag 1440
agaaggaaag gggtgatcca gaacggcaga gagagatgga agaagaggag gatgaggatg 1500
aggatgagga tgaagatgag gatgaacggc agttactggg agaatttgag aaggaactgg 1560
aagggatcct gcttccgtca gaccgagacc ggctccgttc ggaggtgaag gctggcatgg 1620
agcgggaact ggagaacatc atccaggaga cagagaaaga gctggaccca gatgggctga 1680
agaaggagtc agagcgggat cgggcaatgc tggctctcac atccactctc aacaaactca 1740
tcaaaagact ggaggaaaaa cagagtccag agctggtgaa gaagcacaag aaaaagaggg 1800
ttgtccccaa aaagcctccc ccatcacccc aacctacaga ggaggatcct gagcacagag 1860
tccgggtccg ggtcaccaag ctccgtctcg gaggccctaa tcaggatctg actgtcctcg 1920
agatgaaacg ggaaaaccca cagctgaaac aaatcgaggg gctggtgaag gagctgctgg 1980
agagggaggg actcacagct gcagggaaaa ttgagatcaa aattgtccgc ccatgggctg 2040
aagggactga agagggtgca cgttggctga ctgatgagga cacgagaaac ctcaaggaga 2100
tcttcttcaa tatcttggtg ccgggagctg aagaggccca gaaggaacgc cagcggcaga 2160
aagagctgga gagcaattac cgccgggtgt gggctctcc aggtggggag ggcacagggg 2220
acctggacga atttgacttc tgagaccaac actacacttg accccttcacg gaatccagac 2280
tcttcctgga ctggcttgcc tcctccccac ctccccaccc tggaacccct gagggccaaa 2340
cagcagagtg gagctgagct gtggacctct cggcaactc tgtgggtgtg ggggccctgg 2400
gtgaatgctg ctgcccctgc tggcagccac cttgagacct caccgggcct gtgatatttg 2460
ctctcctgaa ctctcactca atcctcttcc tctcctctgt ggcttttcct gttattgtcc 2520
cctaatgata ggatattccc tgctgcctac ctggagattc agtaggatct tttgagtgga 2580
ggtgggtaga gagagcaagg agggcaggac acttagcagg cactgagcaa gcaggccccc 2640
acctgcccctt agtgatgttt ggagtcgttt taccctcttc tattgaattg ccttgggatt 2700
tccttctccc tttccctgcc caccctgtcc cctacaattt gtgcttctga gttgaggagc 2760
cttcacctct gttgctgagg aaatggtaga atgctgccta tcacctccag cacaatccca 2820
gcgaaaaagg tgtgaagcac ccaccatgtt cttgaacaat caggtttcta aataaacacc 2880
ttgggcccat caaaaaaaaa aaaaaa 2906
```

[Continued on next page]

(54) Title: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH

(57) Abstract: Various embodiments of the invention provide human proteins associated with cell growth, differentiation, and death (CGDD) and polynucleotides which identify and encode CGDD. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of CGDD.

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) **Date of publication of the international search report:**
8 May 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C07K 14/435; C12N 15/12; A61K 38/17
US CL : 530/350; 536/23.5;514/12
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)   U.S. : 530/350; 536/23.5;514/12
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)   Compugen, SEQ ID NO: 1

C.	DOCUMENTS CONSIDERED TO BE RELEVANT	
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	YU, W. et al, Large-scale concatenation of cDNA sequencing, Genome Research, 1997, Vol. 7, pages 353-358, see entire document and attached alignment.	3, 4 ---------- 1, 2, 17, 18, and 56

☐ Further documents are listed in the continuation of Box C.   ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"B"	earlier application or patent published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 February 2003 (26.02.2003)	19 MAR 2003
Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	James Martinell
Facsimile No. (703)305-3230	Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/25465

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐   Claim Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐   Claim Nos.:
   because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐   Claim Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐   As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐   As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐   As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐   No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest   ☐   The additional search fees were accompanied by the applicant's protest.

                       ☐   No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1)) (July 1998)